Adv. Appl. Prob. **51,** 236–267 (2019)
*doi:*10.1017/apr.2019.10

SINGULAR VECTOR DISTRIBUTION OF SAMPLE COVARIANCE MATRICES

XIUCAI DING,* *University of Toronto*

Abstract

We consider a class of sample covariance matrices of the form $Q = TXX^*T^*$, where $X = (x_{ij})$ is an $M \times N$ rectangular matrix consisting of independent and identically distributed entries, and T is a deterministic matrix such that T^*T is diagonal. Assuming that M is comparable to N, we prove that the distribution of the components of the right singular vectors close to the edge singular values agrees with that of Gaussian ensembles provided the first two moments of x_{ij} coincide with the Gaussian random variables. For the right singular vectors associated with the bulk singular values, the same conclusion holds if the first four moments of x_{ij} match those of the Gaussian random variables. Similar results hold for the left singular vectors if we further assume that T is diagonal.

Keywords: Random matrix theory; singular vector distribution; deformed Marcenko–Pastur law

2010 Mathematics Subject Classification: Primary 15B52
Secondary 82B44

1. Introduction

In the analysis of multivariate data, a large collection of statistical methods, including principal component analysis, regression analysis, and clustering analysis, require the knowledge of covariance matrices [11]. The advance of data acquisition and storage has led to datasets for which the sample size N and the number of variables M are both large. This high dimensionality cannot be handled using the classical statistical theory.

For applications involving large-dimensional covariance matrices, it is important to understand the local behavior of the the singular values and vectors. Assuming that M is comparable to N, the spectral analysis of the singular values has attracted considerable interest since the seminal work of Marcenko and Pastur [30]. Since then, numerous researchers have contributed to weakening the conditions on matrix entries as well as extending the class of matrices for which the empirical spectral distributions (ESDs) have nonrandom limits. For a detailed review, we refer the reader to the monograph [2]. Besides the ESDs of the singular values, the limiting distributions of the extreme singular values were analysed in a collection of celebrated papers. The results were first proved for the Wishart matrix (i.e. sample covariance matrices obtained from a data matrix consisting of independent and identically distributed (i.i.d.) centered real or complex Gaussian entries) in [23] and [38]; they were later proved for matrices with entries

Received 28 September 2017; revision received 17 September 2018.
The supplementary material for this article can be found at http://doi.org/10.1017/apr.2019.10.
* Postal address: Department of Statistical Sciences, University of Toronto, Sidney Smith Hall, 100 St. George Street, Toronto, ON M5S 3G3, Canada.
Email address: xiucai.ding@mail.utoronto.ca

satisfying arbitrary subexponential distributions in [5], [32], and [33]. More recently, the weakest moment condition was given in [16].

Less is known however for the singular vectors. Therefore, recent research on the limiting behavior of singular vectors has attracted considerable interest among mathematicians and statisticians. Silverstein first derived limit theorems for the eigenvectors of covariance matrices [34]; later, the results were proved for a general class of covariance matrices [3]. The delocalization property for the eigenvectors were shown in [8] and [33]. The universal properties of the eigenvectors of covariance matrices were analysed in [8], [9], [27], and [37]. For a recent survey of the results, we refer the reader to [31]. In this paper we prove the universality for the distribution of the singular vectors for a general class of covariance matrices of the form $Q = TXX^*T^*$, where T is a deterministic matrix such that T^*T is diagonal.

The covariance matrix Q contains a general class of covariance structures and random matrix models [8, Section 1.2]. The singular values analysis of Q has attracted considerable attention; see, for example, the limiting spectral distribution and Stieltjes transform derived in [35], the Tracy–Widom asymptotics of the extreme eigenvalues proved in [5], [17], [26], and [28], and the anisotropic local law proposed in [26]. It is notable that, in general, Q contains the spiked covariance matrices [4], [6], [7], [8], [23]. In such models, the ESD of Q still satisfies the Marcenko–Pastur (MP) law and some of the eigenvalues of Q will detach from the bulk and become outliers. However, in this paper, we adapt the regularity Assumption 1.2 to rule out the outliers for the purpose of universality discussion. Actually, it was shown in [12] and [25] that the distributions of the outliers are not universal.

In this paper we study the singular vector distribution of Q. We prove the universality for the components of the edge singular vectors by assuming the matching of the first two moments of the matrix entries. We also prove similar results in the bulk, under the stronger assumption that the first four moments of the two ensembles match. Similar results have been proved for Wigner matrices in [24].

1.1. Sample covariance matrices with a general class of populations

We first introduce some notation. Throughout the paper, we will use

$$r = \lim_{N\to\infty} r_N = \lim_{N\to\infty} \frac{N}{M}. \tag{1.1}$$

Let $X = (x_{ij})$ be an $M \times N$ data matrix with centered entries $x_{ij} = N^{-1/2} q_{ij}$, $1 \le i \le M$ and $1 \le j \le N$, where q_{ij} are i.i.d. random variables with unit variance and for all $p \in \mathbb{N}$, there exists a constant C_p such that q_{11} satisfies the condition

$$\mathbb{E}|q_{11}|^p \le C_p. \tag{1.2}$$

We consider the sample covariance matrix $Q = TXX^*T^*$, where T is a deterministic matrix such that T^*T is a positive diagonal matrix. Using the QR factorization [22, Theorem 5.2.1], we find that $T = U\Sigma^{1/2}$, where U is an orthogonal matrix and Σ is a positive diagonal matrix. Define $Y = \Sigma^{1/2}X$ and the singular value decomposition of Y as $Y = \sum_{k=1}^{N\wedge M} \sqrt{\lambda_k}\xi_k\zeta_k^*$, where λ_k, $k = 1, 2, \ldots, N \wedge M$, are the nontrivial eigenvalues of Q, and $\{\xi_k\}_{k=1}^M$ and $\{\zeta_k\}_{k=1}^N$ are orthonormal bases of $\mathbb{R}^M$ and $\mathbb{R}^N$, respectively. First, we observe that

$$X^*T^*TX = Y^*Y = \mathbf{Z}\Lambda_N\mathbf{Z}^*,$$

where the columns of $\mathbf{Z}$ are $\zeta_1, \ldots, \zeta_N$ and Λ_N is a diagonal matrix with entries $\lambda_1, \ldots, \lambda_N$. As a consequence, U will not influence the right singular vectors of Y. For the left singular

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

vectors, we need to further assume that T is diagonal. Hence, we can make the following assumption on T:

$$T \equiv \Sigma^{1/2} = \operatorname{diag}\{\sigma_1^{1/2}, \ldots, \sigma_M^{1/2}\}, \quad \text{with } \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_M > 0. \tag{1.3}$$

We denote the empirical spectral distribution of Σ by

$$\pi := \frac{1}{M}\sum_{i=1}^{M} \delta_{\sigma_i}. \tag{1.4}$$

Suppose that there exists some small positive constant τ such that

$$\tau < \sigma_M \leq \sigma_1 \leq \tau^{-1}, \qquad \tau \leq r \leq \tau^{-1}, \qquad \pi([0, \tau]) \leq 1 - \tau. \tag{1.5}$$

For definiteness, in this paper we focus on the real case, i.e. all the entries x_{ij} are real. However, it is clear that our results and proofs can be applied to the complex case after minor modifications if we assume in addition that Re x_{ij} and Im x_{ij} are independent centered random variables with the same variance. To avoid repetition, we summarize the basic assumptions for future reference.

Assumption 1.1. *We assume that X is an $M \times N$ matrix with centered i.i.d. entries satisfying (1.1) and (1.2). We also assume that T is a deterministic $M \times M$ matrix satisfying (1.3) and (1.5).*

From now on, we let $Y = \Sigma^{1/2} X$ and its singular value decomposition $Y = \sum_{k=1}^{N \wedge M} \sqrt{\lambda_k} \xi_k \zeta_k^*$, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{M \wedge N}$.

1.2. Deformed Marcenko–Pastur law

In this subsection we discuss the empirical spectral distribution of $X^* T^* T X$, where we basically follow the discussion of [26, Section 2.2]. It is well known that if π is a compactly supported probability measure on $\mathbb{R}$, letting $r_N > 0$, then, for any $z \in \mathbb{C}_+$, there is a unique $m \equiv m_N(z) \in \mathbb{C}_+$ satisfying

$$\frac{1}{m} = -z + \frac{1}{r_N}\int \frac{x}{1 + mx}\pi(\mathrm{d}x). \tag{1.6}$$

We refer the reader to [26, Lemma 2.2] and [36, Section 5] for more details. In this paper we define the deterministic function $m \equiv m(z)$ as the unique solution of (1.6) with π defined in (1.4). We define by ρ the probability measure associated with m (i.e. m is the Stieltjes transform of ρ) and call it the asymptotic density of $X^* T^* T X$. Our assumption (1.5) implies that the spectrum of Σ cannot be concentrated at 0; thus, it ensures π is a compactly supported probability measure. Therefore, m and ρ are well defined.

Let $z \in \mathbb{C}_+$. Then $m \equiv m(z)$ can be characterized as the unique solution of the equation

$$z = f(m), \qquad \operatorname{Im} m \geq 0, \quad \text{where } f(x) := -\frac{1}{x} + \frac{1}{r_N}\sum_{i=1}^{M} \frac{\pi(\{\sigma_i\})}{x + \sigma_i^{-1}}. \tag{1.7}$$

The behavior of ρ can be entirely understood by the analysis of f. We summarize the elementary properties of ρ in the following lemma. It can be found in [26, Lemmas 2.4, 2.5, and 2.6].

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

Lemma 1.1. *Define* $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. *Then* f *defined in* (1.7) *is smooth on the* $M+1$ *open intervals of* $\overline{\mathbb{R}}$ *defined through*

$$I_1 := \left(-\sigma_1^{-1}, 0\right), \qquad I_i := \left(-\sigma_i^{-1}, -\sigma_{i-1}^{-1}\right), \quad i = 2, \ldots, M, \qquad I_0 := \frac{\overline{R}}{\bigcup_{i=1}^{M} \overline{I}_i}.$$

We also introduce a multiset $\mathcal{C} \subset \overline{\mathbb{R}}$ *containing the critical points of* f, *using the conventions that a nondegenerate critical point is counted once and a degenerate critical point will be counted twice. In the* $r_N = 1$ *case,* ∞ *is a nondegenerate critical point. With the above notation, the following statememts hold.*

- *We have* $|\mathcal{C} \cap I_0| = |\mathcal{C} \cap I_1| = 1$ *and* $|\mathcal{C} \cap I_i| \in \{0, 2\}$ *for* $i = 2, \ldots, M$. *Therefore,* $|\mathcal{C}| = 2p$, *where, for convenience, we denote by* $x_1 \geq x_2 \geq \ldots \geq x_{2p-1}$ *the* $2p-1$ *critical points in* $I_1 \cup \ldots \cup I_M$ *and by* x_{2p} *the unique critical point in* I_0.
- *Defining* $a_k := f(x_k)$ *we have* $a_1 \geq \ldots \geq a_{2p}$. *Moreover, we have* $x_k = m(a_k)$ *by assuming that* $m(0) := \infty$ *for* $r_N = 1$. *Furthermore, for* $k = 1, \ldots, 2p$, *there exists a constant* C *such that* $0 \leq a_k \leq C$.
- *We have* $\operatorname{supp} \rho \cap (0, \infty) = \left(\bigcup_{k=1}^{p} [a_{2k}, a_{2k-1}]\right) \cap (0, \infty)$.

With the above definitions and properties, we now introduce the key regularity assumption on Σ.

Assumption 1.2. *Fix* $\tau > 0$. *We say that*

1. *the edges* a_k, $k = 1, \ldots, 2p$, *are regular if*

$$a_k \geq \tau, \qquad \min_{l \neq k} |a_k - a_l| \geq \tau, \qquad \min_i |x_k + \sigma_i^{-1}| \geq \tau; \tag{1.8}$$

2. *the bulk components* $k = 1, \ldots, p$ *are regular if, for any fixed* $\tau' > 0$, *there exists a constant* $c \equiv c_{\tau,\tau'}$ *such that the density of* ρ *in* $[a_{2k} + \tau', a_{2k-1} - \tau']$ *is bounded from below by* c.

Remark 1.1. The second condition in (1.8) states that the gap in the spectrum of ρ adjacent to a_k can be well separated when N is sufficiently large. The third condition ensures a square root behavior of ρ in a small neighborhood of a_k. To be specific, consider the right edge of the kth bulk component; by Equation (A.12) of [26], there exists some small constant $c > 0$ such that ρ has the following square root behavior:

$$\rho(x) \sim \sqrt{a_{2k-1} - x}, \qquad x \in [a_{2k-1} - c, a_{2k-1}]. \tag{1.9}$$

As a consequence, it will rule out the outliers. The bulk regularity imposes a lower bound on the density of eigenvalues away from the edges. For examples of matrices Σ verifying the regularity conditions, we refer the reader to [26, Examples 2.8 and 2.9].

1.3. Main results

In this subsection we provide the main results of this paper. We first introduce some notation. Recall that the nontrivial classical eigenvalue locations $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_{M \wedge N}$ of Q are defined as $\int_{\gamma_i}^{\infty} \mathrm{d}\rho = (i - \frac{1}{2})/N$. By Lemma 1.1, there are p bulk components in the spectrum of ρ. For $k = 1, \ldots, p$, we define the classical number of eigenvalues of the kth bulk

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

component through $N_k := N\int_{a_{2k}}^{a_{2k-1}} \mathrm{d}\rho$. When $p \geq 1$, we relabel λ_i and γ_i separately for each bulk component $k = 1, \ldots, p$ by introducing

$$\lambda_{k,i} := \lambda_{i+\sum_{l<k} N_l}, \qquad \gamma_{k,i} := \gamma_{i+\sum_{l<k} N_l} \in (a_{2k}, a_{2k-1}). \tag{1.10}$$

Equivalently, we can characterize $\gamma_{k,i}$ through

$$\int_{\gamma_{k,i}}^{a_{2k-1}} \mathrm{d}\rho = \frac{i-1/2}{N}.$$

In this paper we will use the following assumption for the technical application of the anisotropic local law.

Assumption 1.3. *For $k = 1, 2, \ldots, p$ and $i = 1, 2, \ldots, N_k$, $\gamma_{k,i} \geq \tau$ for some constant $\tau > 0$.*

We define the index sets $\mathcal{I}_1 := \{1, \ldots, M\}$ and $\mathcal{I}_2 := \{M+1, \ldots, M+N\}$, with $\mathcal{I} := \mathcal{I}_1 \cup \mathcal{I}_2$. We will consistently use Latin letters $i, j \in \mathcal{I}_1$, Greek letters $\mu, \nu \in \mathcal{I}_2$, and $s, t \in \mathcal{I}$. Then we label the indices of the matrix according to $X = (X_{i\mu} : i \in \mathcal{I}_1,\ \mu \in \mathcal{I}_2)$. We similarly label the entries of $\xi_k \in \mathbb{R}^{\mathcal{I}_1}$ and $\zeta_k \in \mathbb{R}^{\mathcal{I}_2}$. In the kth, $k = 1, 2, \ldots, p$, bulk component, we rewrite the index of $\lambda_{\alpha'}$ as

$$\alpha' := l + \sum_{t<k} N_t \quad \text{when } \alpha' - \sum_{t<k} N_t < \sum_{t\leq k} N_t - \alpha', \tag{1.11}$$

$$\alpha' := -l + 1 + \sum_{t\leq k} N_t \quad \text{when } \alpha' - \sum_{t<k} N_t > \sum_{t\leq k} N_t - \alpha'. \tag{1.12}$$

In this paper we say that l *is associated with* α'. Note that α' is the index of $\lambda_{k,l}$ before the relabeling of (1.10), and the two cases correspond to the right and left edges, respectively. Our main result on the distribution of the components of the singular vectors near the edge is the following theorem. For any positive integers m, k, some function $\theta : \mathbb{R}^m \to \mathbb{R}$, and $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$, we define

$$\partial^{(k)}\theta(x) = \frac{\partial^k \theta(x)}{\partial x_1^{k_1} \partial x_2^{k_2} \ldots \partial x_m^{k_m}}, \qquad \sum_{i=1}^{m} k_i = k,\ k_1, k_2, \ldots, k_m \geq 0,$$

and $||x||_2$ to be its l_2 norm. Define $Q_G := \Sigma^{1/2} X_G X_G^* \Sigma^{1/2}$, where X_G is GOE (i.e. a random matrix with entries being i.i.d. real standard Gaussian random variables) and Σ satisfies (1.3) and (1.5).

Theorem 1.1. *Suppose that $Q_V = \Sigma^{1/2} X_V X_V^* \Sigma^{1/2}$ satisfies Assumption 1.1. Let $\mathbb{E}^G$ and $\mathbb{E}^V$ denote the expectations with respect to X_G and X_V. Consider the kth, $k = 1, 2, \ldots, p$, bulk component, with l defined in (1.11) or (1.12). Under Assumption 1.2 and 1.3 for any choices of indices $i, j \in \mathcal{I}_1$ and $\mu, \nu \in \mathcal{I}_2$, there exists a $\delta \in (0, 1)$ such that, when $l \leq N_k^{\delta}$, we have*

$$\lim_{N\to\infty} [\mathbb{E}^V - \mathbb{E}^G]\theta(N\xi_{\alpha'}(i)\xi_{\alpha'}(j), N\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)) = 0,$$

where θ is a smooth function in $\mathbb{R}^2$ that satisfies

$$|\partial^{(k)}\theta(x)| \leq C(1 + ||x||_2)^C, \qquad k = 1, 2, 3, \textit{ with some constant } C > 0.$$

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

Theorem 1.2. *Suppose that $Q_V = \Sigma^{1/2} X_V X_V^* \Sigma^{1/2}$ satisfies Assumption 1.1. Consider the k_1th, ..., k_nth, $k_1, \ldots, k_n \in \{1, 2, \ldots, p\}$, $n \le p$, bulk components for l_{k_i} defined in (1.11) or (1.12) associated with the k_ith, $i = 1, 2, \ldots, n$, bulk component. Under Assumptions 1.2 and 1.3 for any choices of indices $i, j \in \mathcal{I}_1$ and $\mu, \nu \in \mathcal{I}_2$, there exists a $\delta \in (0, 1)$ such that, when $l_{k_i} \le N_{k_i}^{\delta}$, where l_{k_i} is associated with α'_{k_i}, $i = 1, 2, \ldots, n$, we have*

$$\lim_{N\to\infty} [\mathbb{E}^V - \mathbb{E}^G]\theta\big(N\xi_{\alpha'_{k_1}}(i)\xi_{\alpha'_{k_1}}(j), N\zeta_{\alpha'_{k_1}}(\mu)\zeta_{\alpha'_{k_1}}(\nu), \ldots, N\xi_{\alpha'_{k_n}}(i)\xi_{\alpha'_{k_n}}(j), N\zeta_{\alpha'_{k_n}}(\mu)\zeta_{\alpha'_{k_n}}(\nu)\big) = 0,$$

where θ is a smooth function in $\mathbb{R}^{2n}$ that satisfies

$$|\partial^{(k)}\theta(x)| \le C(1 + ||x||_2)^C, \qquad k = 1, 2, 3, \textit{ with some constant } C > 0.$$

Remark 1.2. The results in Theorems 1.1 and 1.2 can be easily extended to a general form containing more entries of the singular vectors using a general form of the Green function comparison argument. For example, to extend Theorem 1.1, we consider the kth bulk component and choose any positive integer s. Under Assumptions 1.2 and 1.3 for any choices of indices $i_1, j_1, \ldots, i_s, j_s \in \mathcal{I}_1$ and $\mu_1, \nu_1, \ldots, \mu_s, \nu_s \in \mathcal{I}_2$ for the corresponding l_i, $i = 1, 2, \ldots, s$, defined in (1.11) or (1.12), there exists some $0 < \delta < 1$ with $0 < \max_{1\le i\le s}\{l_i\} \le N_k^{\delta}$, such that

$$\lim_{N\to\infty} \big[\mathbb{E}^V - \mathbb{E}^G\big]\theta(N\xi_{\alpha'_1}(i_1)\xi_{\alpha'_1}(j_1), N\zeta_{\alpha'_1}(\mu_1)\zeta_{\alpha'_1}(\nu_1), \ldots, N\xi_{\alpha'_s}(i_s)\xi_{\alpha'_s}(j_s), N\zeta_{\alpha'_s}(\mu_s)\zeta_{\alpha'_s}(\nu_s)) = 0, \tag{1.13}$$

where $\theta \in \mathbb{R}^{2s}$ is a smooth function satisfying $|\partial^{(k)}\theta(x)| \le C(1 + ||x||_2)^C$, $k = 1, 2, 3$, with some constant $C > 0$. Similarly, we can extend Theorem 1.2 to contain more entries of singular vectors.

Recall (1.10), and define $\varpi_k := (|f''(x_k)|/2)^{1/3}$, $k = 1, 2, \ldots, 2p$. Then, for any positive integer h, we define

$$\boldsymbol{q}_{2k-1,h} := \frac{N^{2/3}}{\varpi_{2k-1}}(\lambda_{k,h} - a_{2k-1}), \qquad \boldsymbol{q}_{2k,h} := -\frac{N^{2/3}}{\varpi_{2k}}(\lambda_{k,N_k-h+1} - a_{2k}).$$

Consider a smooth function $\theta \in \mathbb{R}$ whose third derivative $\theta^{(3)}$ satisfies $|\theta^{(3)}(x)| \le C(1 + |x|)^C$ for some constant $C > 0$. Then, by [26, Theorem 3.18], we have

$$\lim_{N\to\infty} [\mathbb{E}^V - \mathbb{E}^G]\theta(\boldsymbol{q}_{k,h}) = 0. \tag{1.14}$$

Together with Theorem 1.1, we have the following corollary, which is an analogy of [24, Theorem 1.6]. Let $t = 2k - 1$ if α' is as given in (1.11) and $2k$ if α' is as given in (1.12).

Corollary 1.1. *Under the assumptions of Theorem 1.1, for some positive integer h, we have*

$$\lim_{N\to\infty} [\mathbb{E}^V - \mathbb{E}^G]\theta\big(\boldsymbol{q}_{t,h}, N\xi_{\alpha'}(i)\xi_{\alpha'}(j), N\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)\big) = 0,$$

where $\theta \in \mathbb{R}^3$ satisfies

$$|\partial^{(k)}\theta(x)| \le C(1 + ||x||_2)^C, \qquad k = 1, 2, 3, \textit{ with some constant } C > 0. \tag{1.15}$$

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

Corollary 1.1 can be extended to a general form for several bulk components. Let $t_i = 2k_i - 1$ if α'_{k_i} is as given in (1.11) and $2k_i$ if α'_{k_i} is as given in (1.12).

Corollary 1.2. *Under the assumptions of Theorem 1.2, for some positive integer h, we have*

$$\lim_{N\to\infty} [\mathbb{E}^V - \mathbb{E}^G]\theta\big(\boldsymbol{q}_{t_1,h}, N\xi_{\alpha'_{k_1}}(i)\xi_{\alpha'_{k_1}}(j), N\zeta_{\alpha'_{k_1}}(\mu)\zeta_{\alpha'_{k_1}}(\nu), \ldots, \boldsymbol{q}_{t_n,h}, N\xi_{\alpha'_{k_n}}(i)\xi_{\alpha'_{k_n}}(j), \times N\zeta_{\alpha'_{k_n}}(\mu)\zeta_{\alpha'_{k_n}}(\nu)\big) = 0,$$

where $\theta \in \mathbb{R}^{3n}$ is a smooth function that satisfies

$$|\partial^{(k)}\theta(x)| \le C(1 + ||x||_2)^C, \qquad k = 1, 2, 3, \textit{ with some arbitrary } C > 0.$$

Remark 1.3. (i) Similarly to (1.13), the results in Corollaries 1.1 and 1.2 can be easily extended to a general form containing more entries of the singular vectors. For example, to extend Corollary 1.1, we choose any positive integers s and $h_1, \ldots, h_s$. Under Assumptions 1.2 and 1.3 for any choices of indices $i_1, j_1, \ldots, i_s, j_s \in \mathcal{I}_1$ and $\mu_1, \nu_1, \ldots, \mu_s, \nu_s \in \mathcal{I}_2$, for the corresponding l_i, $i = 1, 2, \ldots, s$, defined in (1.11) or (1.12), there exists some $0 < \delta < 1$ with $\max_{1\le i\le s}\{l_i\} \le N_k^{\delta}$, such that

$$\lim_{N\to\infty} [\mathbb{E}^V - \mathbb{E}^G]\theta\big(\boldsymbol{q}_{t_1,h_1}, N\xi_{\alpha'_1}(i_1)\xi_{\alpha'_1}(j_1), \zeta_{\alpha'_1}(\mu_1)\zeta_{\alpha'_1}(\nu_1), \ldots, \boldsymbol{q}_{t_s,h_s}, N\xi_{\alpha'_s}(i_s)\xi_{\alpha'_s}(j_s), N\zeta_{\alpha'_s}(\mu_s)\zeta_{\alpha'_s}(\nu_s)\big) = 0.$$

where the smooth function $\theta \in \mathbb{R}^{3s}$ satisfies $|\partial^{(k)}\theta(x)| \le C(1 + ||x||_2)^C$, $k = 1, 2, 3$, for some constant C.

(ii) Theorems 1.1 and 1.2, and Corollaries 1.1 and 1.2 still hold for the complex case, where the moment matching condition is replaced by

$$\mathbb{E}^G \bar{x}_{ij}^v x_{ij}^u = \mathbb{E}^V \bar{x}_{ij}^v x_{ij}^u, \qquad 0 \le v + u \le 2.$$

(iii) All the above theorems and corollaries are stronger than their counterparts from [24] because they hold much further into the bulk components. For instance, in the counterpart of Theorem 1.1, which is [24, Theorem 1.6], the universality was established under the assumption that $l \le (\log N)^{C \log\log N}$.

In the bulks, similar results hold under the stronger assumption that the first four moments of the matrix entries match those of Gaussian ensembles.

Theorem 1.3. *Suppose that $Q_V = \Sigma^{1/2} X_V X_V^* \Sigma^{1/2}$ satisfies Assumption 1.1. Assume that the third and fourth moments of X_V agree with those of X_G and consider the kth, $k = 1, 2, \ldots, p$ bulk component, with l defined in (1.11) or (1.12). Under Assumptions 1.2 and 1.3 for any choices of indices $i, j \in \mathcal{I}_1$ and $\mu, \nu \in \mathcal{I}_2$, there exists a small $\delta \in (0, 1)$ such that, when $\delta N_k \le l \le (1 - \delta)N_k$, we have*

$$\lim_{N\to\infty} [\mathbb{E}^V - \mathbb{E}^G]\theta(N\xi_{\alpha'}(i)\xi_{\alpha'}(j), N\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)) = 0,$$

where θ is a smooth function in $\mathbb{R}^2$ that satisfies

$$|\partial^{(k)}\theta(x)| \le C(1 + ||x||_2)^C, \qquad k = 1, 2, 3, 4, 5, \textit{ with some constant } C > 0.$$

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

Theorem 1.4. *Suppose that $Q_V = \Sigma^{1/2} X_V X_V^* \Sigma^{1/2}$ satisfies Assumption 1.1. Assume that the third and fourth moments of X_V agree with those of X_G, and consider the k_1th, ..., k_nth, $k_1, \ldots, k_n \in \{1, 2, \ldots, p\}$, $n \le p$, bulks for l_{k_i} defined in (1.11) or (1.12) associated with the k_ith, $i = 1, 2, \ldots, n$, bulk component. Under Assumptions 1.2 and 1.3 for any choices of indices $i, j \in \mathcal{I}_1$ and $\mu, \nu \in \mathcal{I}_2$, there exists a $\delta \in (0, 1)$ such that, when $\delta N_{k_i} \le l_{k_i} \le (1-\delta) N_{k_i}$, $i = 1, 2, \ldots, n$, we have*

$$\lim_{N\to\infty} [\mathbb{E}^V - \mathbb{E}^G]\theta\big(N\xi_{\alpha'_{k_1}}(i)\xi_{\alpha'_{k_1}}(j), N\zeta_{\alpha'_{k_1}}(\mu)\zeta_{\alpha'_{k_1}}(\nu), \ldots, N\xi_{\alpha'_{k_n}}(i)\xi_{\alpha'_{k_n}}(j), N\zeta_{\alpha'_{k_n}}(\mu)\zeta_{\alpha'_{k_n}}(\nu)\big)$$
$$= 0,$$

where θ is a smooth function in $\mathbb{R}^{2n}$ that satisfies

$$|\partial^{(k)}\theta(x)| \le C(1 + ||x||_2)^C, \qquad k = 1, 2, 3, 4, 5, \text{ with some constant } C > 0.$$

Remark 1.4. (i) Similarly to Corollaries 1.1 and 1.2 and Remark 1.3(i), we can extend the results to the joint distribution containing singular values. We take the extension of Theorem 1.3 as an example. By Assumption 1.2(ii), in the bulk, we have $\int_{\lambda_{\alpha'}}^{\gamma_{\alpha'}} \mathrm{d}\rho = 1/N + o(N^{-1})$. Using a similar Dyson Brownian motion argument as in [33], combining with Theorem 1.3, we have

$$\lim_{N\to\infty} [\mathbb{E}^V - \mathbb{E}^G]\theta(\boldsymbol{p}_{\alpha'}, N\xi_{\alpha'}(i)\xi_{\alpha'}(j), N\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)) = 0,$$

where $\boldsymbol{p}_{\alpha'}$ is defined as

$$\boldsymbol{p}_{\alpha'} := \rho(\gamma_{\alpha'})N(\lambda_{\alpha'} - \gamma_{\alpha'}),$$

and $\theta \in \mathbb{R}^3$ satisfies

$$|\partial^{(k)}\theta(x)| \le C(1 + ||x||_2)^C, \qquad k = 1, 2, 3, 4, 5, \text{ with some constant } C > 0.$$

(ii) Theorems 1.3 and 1.4 still hold for the complex case, where the moment matching condition is replaced by

$$\mathbb{E}^G \bar{x}_{ij}^v x_{ij}^u = \mathbb{E}^V \bar{x}_{ij}^v x_{ij}^u, \qquad 0 \le v + u \le 4.$$

1.4. Remarks on applications to statistics

In this subsection we give a few remarks on possible applications to statistics and machine learning. First, our results show that, under Assumptions 1.1, 1.2, and 1.3, the distributions of the right singular vectors, i.e. entries of principal components, are independent of the laws of x_{ij}. Hence, we can extend the statistical analysis relying on Gaussian or sub-Gaussian assumptions to general distributions. For instance, in the problem of classification, assuming that $Y = (y_i)$ and each y_i has the same covariance structure but may have different means, i.e. $\mathbb{E}y_i = \mu_k$, $i = 1, 2, \ldots, N$, $k = 1, 2, \ldots, K$, where K is a fixed constant. We are interested in classifying the samples y_i into K clusters. In the classical framework, researchers use the matrix ΛV to classify the samples y_i, where $\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_K\}$ and $V = (\zeta_1, \ldots, \zeta_K)$ (recall that λ_i and ζ_i are the singular values and right singular vectors of Y). Existing statistical analysis needs the sub-Gaussian assumption [29]. In this sense, our result, especially Remark 1.4, can be used to generalize such results.

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

Next, our results can be used to do statistical inference. It is notable that, in general, the distribution of the singular vectors of the sample covariance matrix $Q = TXX^*T^*$ is unknown, even for the Gaussian case. However, when T is a scalar matrix (i.e. $T = cI,\ c > 0$), Bourgade and Yau [10, Appendix C] showed that the entries of the singular vectors are asymptotically normally distributed. Hence, our universality results imply that, under Assumptions 1.1, 1.2, and 1.3, when T is conformal (i.e. $T^*T = cI,\ c > 0$), the entries of the right singular vectors are asymptotically normally distributed. Therefore, this can be used to test the null hypothesis:

($\mathbf{H}_0$) T is a conformal matrix.

The statistical testing problem ($\mathbf{H}_0$) contains a rich class of hypothesis tests. For instance, when $T = I$, it reduces to the sphericity test and when $c = 1$, it reduces to testing whether the covariance matrix of X is orthogonal [40].

To illustrate how our results can be used to test ($\mathbf{H}_0$), we assume that $c = 1$ in the following discussion. Under ($\mathbf{H}_0$), denote the QR factorization of T to be $T = UI$, the right singular vector of TX is the same as X, $\zeta_k,\ k = 1, 2, \ldots, N$. Using [10, Corollary 1.3], we find that, for $i, k = 1, 2, \ldots, N$,

$$\sqrt{N}\zeta_k(i) \to \mathcal{N}, \tag{1.16}$$

where $\mathcal{N}$ is a standard Gaussian random variable. In detail, we can take the following steps to test whether ($\mathbf{H}_0$) holds.

1. Randomly choose two index sets $R_1, R_2 \subset \{1, 2, \ldots, N\}$ with $|R_i| = O(1),\ i = 1, 2$.
2. Use the bootstrapping method to sample the columns of Q and obtain a sequence of $M \times N$ matrices $Q_j,\ j = 1, 2, \ldots, K$.
3. Select $\zeta_k^j(i),\ k \in R_1, i \in R_2$, from $Q_j,\ j = 1, 2, \ldots, K$. Use the classic normality test, for instance, the Shapiro–Wilk test, to check whether (1.16) holds for the above samples. Let A be the number of samples which cannot be rejected by the classic normality test.
4. Given some pre-chosen significant level α, reject $\mathbf{H}_0$ if $A/|R_1||R_2| < 1 - \alpha$.

Another important piece of information from our result is that the singular vectors are completely delocalized. This property can be applied to the problem of low rank matrix denoising [13], i.e.

$$\hat{S} = TX + S,$$

where S is a deterministic low rank matrix. Consider that S is of rank one, and assume that the left singular vector u of S is $\mathbf{e}_1 = (1, 0, \ldots, 0) \in \mathbb{R}^M$. Using the completely delocalized result, it can be shown that $\tilde{u}_1$, the first left singular vector of $\hat{S}$ has the same sparse structure as that of u, i.e.

$$\tilde{u}_1(1) = O(1), \qquad \tilde{u}_1(i) = O(M^{-1/2}), \quad i \neq 1,$$

hold with high probability. Thus, to estimate the singular vectors of S, we need only carry out singular value decomposition on a block matrix of $\hat{S}$. For more details, we refer the reader to [13, Section 2.1].

Furthermore, delocalization of singular vectors is important in machine learning, especially the perturbation analysis of a singular subspace [1], [15], [21], [20], [41]. In these problems, researchers are interested in bounding the difference between the sample singular vectors and those of T. The Davis–Kahan $\sin\theta$ theorem is often used to bound the l_2 distance.

However, in many applications, for instance, the wireless sensor network localization [21] and multidimesional scaling [15], people are usually interested in bounding the l_∞ distance. Denote the right singular vectors of T by v_i and recall that the ζ_i are the right singular vectors of Y. We aim to bound

$$||v_i - \zeta_i||_\infty.$$

To obtain such a bound, an important step is to show the delocalization (i.e. incoherence) of the singular vectors [1], [15], [41]. Hence, our results in this paper can provide the crucial ingredients for such applications.

This paper is organized as follows. In Section 2 we introduce some notation and tools that will be used in the proofs. In Section 3 we prove the singular vector distribution near the edge. In Section 4 we prove the distribution within the bulks. The Green function comparison arguments are mainly discussed in Section 3.2 and Lemma 4.5. The proof of Lemma 3.4 is given in the supplementary material [14] to this paper.

Conventions. We always use C to denote a generic large positive constant, whose value may change from one line to the next. Similarly, we use ε to denote a generic small positive constant. For two quantities a_N and b_N depending on N, the notation $a_N = O(b_N)$ means that $|a_N| \le C|b_N|$ for some positive constant $C > 0$, and $a_N = o(b_N)$ means that $|a_N| \le c_N|b_N|$ for some positive constants $c_N \to 0$ as $N \to \infty$. We also use the notation $a_N \sim b_N$ if $a_N = O(b_N)$ and $b_N = O(a_N)$. We write the identity matrix $I_{n\times n}$ as 1 or I when there is no confusion about the dimension.

2. Notation and tools

In this section we introduce some notation and tools which will be used in this paper. Throughout the paper, we always use ε_1 to denote a small constant and D_1 to denote a large constant. Recall that the ESD of an $N \times N$ symmetric matrix H is defined as

$$F_H^{(N)}(\lambda) := \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\{\lambda_i(H)\le\lambda\}},$$

and its Stieltjes transform is defined as

$$m_H = \int \frac{1}{x-z}\, \mathrm{d}F_H^{(N)}(x), \qquad z = E + \mathrm{i}\eta \in \mathbb{C}_+.$$

For some small constant $\tau > 0$, we define the typical domain for $z = E + \mathrm{i}\eta$ as

$$\boldsymbol{D}(\tau) = \{z \in \mathbb{C}_+ : |E| \le \tau^{-1},\ N^{-1+\tau} \le \eta \le \tau^{-1}\}. \tag{2.1}$$

It was shown in [13], [16], [26], and [39] that the linearizing block matrix is quite useful in dealing with rectangular matrices.

Definition 2.1. For $z \in \mathbb{C}_+$, we define the $(N+M)\times(N+M)$ self-adjoint matrix

$$H \equiv H(X, \Sigma) := \begin{pmatrix} -zI & z^{1/2}Y \\ z^{1/2}Y^* & -zI \end{pmatrix}, \tag{2.2}$$

and

$$G \equiv G(X, z) := H^{-1}. \tag{2.3}$$

By Schur's complement, it is easy to check that

$$G=\begin{pmatrix}\mathcal{G}_1(z) & z^{-1/2}\mathcal{G}_1(z)Y\\ z^{-1/2}Y^*\mathcal{G}_1(z) & z^{-1}Y^*\mathcal{G}_1(z)Y-z^{-1}I\end{pmatrix}$$
$$=\begin{pmatrix}z^{-1}Y\mathcal{G}_2(z)Y^*-z^{-1}I & z^{-1/2}Y\mathcal{G}_2(z)\\ z^{-1/2}\mathcal{G}_2(z)Y^* & \mathcal{G}_2(z)\end{pmatrix}, \tag{2.4}$$

where

$$\mathcal{G}_1(z):=(YY^*-z)^{-1},\qquad \mathcal{G}_2(z):=(Y^*Y-z)^{-1},\qquad z=E+i\eta\in\mathbb{C}_+.$$

Thus, a control of G directly yields controls of $(YY^*-z)^{-1}$ and $(Y^*Y-z)^{-1}$. Moreover, we have

$$m_1(z)=\frac{1}{M}\sum_{i\in\mathcal{I}_1}G_{ii},\qquad m_2(z)=\frac{1}{N}\sum_{\mu\in\mathcal{I}_2}G_{\mu\mu}. \tag{2.5}$$

Recall that $Y=\sum_{i=1}^{M\wedge N}\sqrt{\lambda_k}\xi_k\zeta_k^*,\ \xi_k\in\mathbb{R}^{\mathcal{I}_1},\ \zeta_k\in\mathbb{R}^{\mathcal{I}_2}$. By (2.4), we have

$$G(z)=\sum_{k=1}^{M\wedge N}\frac{1}{\lambda_k-z}\begin{pmatrix}\xi_k\xi_k^* & z^{-1/2}\sqrt{\lambda_k}\xi_k\zeta_k^*\\ z^{-1/2}\sqrt{\lambda_k}\zeta_k\xi_k^* & \zeta_k\zeta_k^*\end{pmatrix}. \tag{2.6}$$

Define

$$\Psi(z):=\sqrt{\frac{\operatorname{Im} m(z)}{N\eta}}+\frac{1}{N\eta},\qquad \underline{\Sigma}_o:=\begin{pmatrix}\Sigma & 0\\ 0 & I\end{pmatrix},\qquad \underline{\Sigma}:=\begin{pmatrix}z^{-1/2}\Sigma^{1/2} & 0\\ 0 & I\end{pmatrix}. \tag{2.7}$$

Definition 2.2. For $z\in\mathbb{C}_+$, we define the $\mathcal{I}\times\mathcal{I}$ matrix

$$\Pi(z):=\begin{pmatrix}-z^{-1}(1+m(z)\Sigma)^{-1} & 0\\ 0 & m(z)\end{pmatrix}. \tag{2.8}$$

We will see later from Lemma 2.1 that $G(z)$ converges to $\Pi(z)$ in probability.

Remark 2.1. In [26, Definition 3.2], the linearizing block matrix is defined as

$$H_o:=\begin{pmatrix}-\Sigma^{-1} & X\\ X^* & -zI\end{pmatrix}. \tag{2.9}$$

It is easy to check the following relation between (2.2) and (2.9):

$$H=\begin{pmatrix}z^{1/2}\Sigma^{1/2} & 0\\ 0 & I\end{pmatrix}H_o\begin{pmatrix}z^{1/2}\Sigma^{1/2} & 0\\ 0 & I\end{pmatrix}. \tag{2.10}$$

In [26, Definition 3.3], the deterministic convergent limit of H_o^{-1} is

$$\Pi_o(z)=\begin{pmatrix}-\Sigma(1+m(z)\Sigma)^{-1} & 0\\ 0 & m(z)\end{pmatrix}. \tag{2.11}$$

Therefore, by (2.10), we can get a similar relation between (2.8) and (2.11):

$$\Pi(z)=\begin{pmatrix}z^{-1/2}\Sigma^{-1/2} & 0\\ 0 & I\end{pmatrix}\Pi_o(z)\begin{pmatrix}z^{-1/2}\Sigma^{-1/2} & 0\\ 0 & I\end{pmatrix}. \tag{2.12}$$

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

Definition 2.3. We introduce the notation $X^{(\mathbb{T})}$ to represent the $M \times (N - |\mathbb{T}|)$ minor of X by deleting the ith, $i \in \mathbb{T}$, columns of X. For convenience, $(\{i\})$ will be abbreviated to (i). We will continue to use the matrix indices of X for $X^{(\mathbb{T})}$, that is, $X_{ij}^{(\mathbb{T})} = \mathbf{1}(j \notin \mathbb{T})X_{ij}$. Let

$$Y^{(\mathbb{T})} = \Sigma^{1/2}X^{(\mathbb{T})}, \qquad \mathcal{G}_1^{(\mathbb{T})} = (Y^{(\mathbb{T})}Y^{(\mathbb{T})^*} - zI)^{-1}, \qquad \mathcal{G}_2^{(\mathbb{T})} = (Y^{(\mathbb{T})^*}Y^{(\mathbb{T})} - zI)^{-1}.$$

Consequently, $m_1^{(\mathbb{T})}(z) = M^{-1}\operatorname{Tr}\mathcal{G}_1^{(\mathbb{T})}(z)$ and $m_2^{(\mathbb{T})}(z) = N^{-1}\operatorname{Tr}\mathcal{G}_2^{(\mathbb{T})}(z)$.

Our key ingredient is the anisotropic local law derived by Knowles and Yin [26].

Lemma 2.1. *Fix $\tau > 0$. Assume that (1.1), (1.2), and (1.5) hold. Moreover, suppose that every edge $k = 1, \dots, 2p$ satisfies $a_k \geq \tau$ and that every bulk component $k = 1, \dots, p$ is regular in the sense of Assumption 1.2. Then, for all $z \in \boldsymbol{D}(\tau)$ and any unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{M+N}$, there exist some small constant $\varepsilon_1 > 0$ and large constant $D_1 > 0$ such that, when N is large enough, with probability $1 - N^{-D_1}$, we have*

$$| < \mathbf{u}, \underline{\Sigma}^{-1}(G(z) - \Pi(z))\underline{\Sigma}^{-1}\mathbf{v} > | \leq N^{\varepsilon_1}\Psi(z) \tag{2.13}$$

and

$$|m_2(z) - m(z)| \leq N^{\varepsilon_1}\Psi(z). \tag{2.14}$$

Proof. Equation (2.14) was proved in [26, Equation (3.11)]. We need only prove (2.13). By (2.10), we have

$$G_o(z) = \begin{pmatrix} z^{1/2}\Sigma^{1/2} & 0 \\ 0 & I \end{pmatrix} G(z) \begin{pmatrix} z^{1/2}\Sigma^{1/2} & 0 \\ 0 & I \end{pmatrix}. \tag{2.15}$$

By [26, Theorem 3.6], with probability $1 - N^{-D_1}$, we have

$$| < \mathbf{u}, \underline{\Sigma}_o^{-1}(G_o(z) - \Pi_o(z))\underline{\Sigma}_o^{-1}\mathbf{v} > | \leq N^{\varepsilon_1}\Psi(z). \tag{2.16}$$

Therefore, by (2.12), (2.15), and (2.16), we conclude our proof. □

It is easy to derive the following corollary from Lemma 2.1.

Corollary 2.1. *Under the assumptions of Lemma 2.1, with probability $1 - N^{-D_1}$, we have*

$$|\langle v, (\mathcal{G}_2(z) - m(z))v\rangle| \leq N^{\varepsilon_1}\Psi(z), \qquad |\langle u, (\mathcal{G}_1(z) + z^{-1}(1 + m(z)\Sigma)^{-1})u\rangle| \leq N^{\varepsilon_1}\Psi(z), \tag{2.17}$$

where v and u are unit vectors in $\mathbb{R}^N$ and $\mathbb{R}^M$, respectively.

We use the following lemma to characterize the rigidity of the eigenvalues within each bulk component, which can be found in [26, Theorem 3.12].

Lemma 2.2. *Fix $\tau > 0$. Assume that (1.1), (1.2), and (1.5) hold. Moreover, suppose that every edge $k = 1, \dots, 2p$ satisfies $a_k \geq \tau$ and that every bulk component $k = 1, \dots, p$ is regular in the sense of Assumption 1.2. Recall that N_k is the number of eigenvalues within each bulk. Then, for $i = 1, \dots, N_k$ satisfying $\gamma_{k,i} \geq \tau$ and $k = 1, \dots, p$, with probability $1 - N^{-D_1}$, we have*

$$|\lambda_{k,i} - \gamma_{k,i}| \leq (i \wedge (N_k + 1 - i))^{-1/3}N^{-2/3+\varepsilon_1}. \tag{2.18}$$

Within the bulk, we have a stronger result. For small $\tau' > 0$, define

$$\boldsymbol{D}_k^b := \big\{z \in \boldsymbol{D}(\tau) \colon E \in [a_{2k} + \tau', a_{2k-1} - \tau']\big\}, \qquad k = 1, 2, \dots, p, \tag{2.19}$$

as the bulk spectral domain. Then [26, Theorem 3.15] gives the following result.

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

Lemma 2.3. *Fix* $\tau, \tau' > 0$. *Assume that (1.1), (1.2), and (1.5) hold and that the bulk component* $k = 1, \ldots, 2p$ *is regular in the sense of Assumption 1.2(ii). Then, for all* $i = 1, \ldots, N_k$ *satisfying* $\gamma_{k,i} \in [a_{2k} + \tau', a_{2k-1} - \tau']$, *(2.13) and (2.14) hold uniformly for all* $z \in \mathbf{D}_k^b$ *and, with probability* $1 - N^{-D_1}$,

$$|\lambda_{k,i} - \gamma_{k,i}| \leq N^{-1+\varepsilon_1}.$$

As discussed in [26, Remark 3.13], Lemmas 2.1 and 2.2 imply complete delocalization of the singular vectors.

Lemma 2.4. *Fix* $\tau > 0$. *Under the assumptions of Lemma 2.1, for any* i *and* μ *such that* $\gamma_i, \gamma_\mu \geq \tau$, *with probability* $1 - N^{-D_1}$, *we have*

$$\max_{i,s_1} |\xi_i(s_1)|^2 + \max_{\mu,s_2} |\zeta_\mu(s_2)|^2 \leq N^{-1+\varepsilon_1}. \tag{2.20}$$

Proof. By (2.17), with probability $1 - N^{-D_1}$, we have $\max\{\operatorname{Im} G_{ii}(z), \ \operatorname{Im} G_{\mu\mu}(z)\} = O(1)$. Choosing $z_0 = E + i\eta_0$ with $\eta_0 = N^{-1+\varepsilon_1}$ and using the spectral decomposition (2.6) yields

$$\sum_{k=1}^{N\wedge M} \frac{\eta_0}{(E-\lambda_k)^2 + \eta_0^2} |\xi_k(i)|^2 = \operatorname{Im} G_{ii}(z_0) = O(1), \tag{2.21}$$

$$\sum_{k=1}^{N\wedge M} \frac{\eta_0}{(E-\lambda_k)^2 + \eta_0^2} |\zeta_k(\mu)|^2 = \operatorname{Im} G_{\mu\mu}(z_0) = O(1), \tag{2.22}$$

with probability $1 - N^{-D_1}$. Choosing $E = \lambda_k$ in (2.21) and (2.22) completes the proof. □

3. Singular vectors near the edges

In this section we prove universality for the distributions of the edge singular vectors of Theorems 1.1 and 1.2, as well as the joint distribution between the singular values and singular vectors of Corollaries 1.1 and 1.2. The main identities on which we will rely are

$$\tilde{G}_{ij} = \sum_{\beta=1}^{M\wedge N} \frac{\eta}{(E-\lambda_\beta)^2 + \eta^2} \xi_\beta(i)\xi_\beta(j), \qquad \tilde{G}_{\mu\nu} = \sum_{\beta=1}^{M\wedge N} \frac{\eta}{(E-\lambda_\beta)^2 + \eta^2} \zeta_\beta(\mu)\zeta_\beta(\nu), \tag{3.1}$$

where $\tilde{G}_{ij}$ and $\tilde{G}_{\mu\nu}$ are defined as

$$\tilde{G}_{ij} := \frac{1}{2i}(G_{ij}(z) - G_{ij}(\bar{z})), \qquad \tilde{G}_{\mu\nu} := \frac{1}{2i}(G_{\mu\nu}(z) - G_{\mu\nu}(\bar{z})).$$

Owing to similarity, we focus our proofs on the right singular vectors. The proofs rely on three main steps.

1. Writing $N\zeta_\beta(\mu)\zeta_\beta(\nu)$ as an integral of $\tilde{G}_{\mu\nu}$ over a random interval with size $O(N^\varepsilon \eta)$, where $\varepsilon > 0$ is a small constant and $\eta = N^{-2/3-\varepsilon_0}$, $\varepsilon_0 > 0$, will be chosen later.
2. Replacing the sharp characteristic function from step (i) with a smooth cutoff function q in terms of the Green function.
3. Using the Green function comparison argument to compare the distribution of the singular vectors between the ensembles X_G and X_V.

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

We will follow the proof strategy of [24, Section 3] and slightly modify the details. Specifically, the choices of random interval in step (i) and the smooth function q in step (ii) are different due to the fact that we have more than one bulk component. The Green function comparison argument is also slightly different as we use the linearization matrix (2.6).

We mainly focus on a single bulk component, first proving the singular vector distribution and then extending the results to singular values. The results containing several bulk components will follow after minor modification. We first prove the following result for the right singular vector.

Lemma 3.1. *Suppose that $Q_V = \Sigma^{1/2} X_V X_V^* \Sigma^{1/2}$ satisfies Assumption 1.1. Let $\mathbb{E}^G, \mathbb{E}^V$ denote the expectations with respect to X_G and X_V. Consider the kth, $k = 1, 2, \ldots, p$, bulk component, with l defined in (1.11) or (1.12), under Assumptions 1.2 and 1.3 for any choices of indices $\mu, \nu \in \mathcal{I}_2$, there exists a $\delta \in (0, 1)$ such that, when $l \le N_k^\delta$, we have*

$$\lim_{N\to\infty} [\mathbb{E}^V - \mathbb{E}^G]\theta(N\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)) = 0,$$

where θ is a smooth function in $\mathbb{R}$ that satisfies

$$|\theta^{(3)}(x)| \le C_1(1+|x|)^{C_1}, \qquad x \in \mathbb{R}, \text{ with some constant } C_1 > 0. \tag{3.2}$$

Near the edges, by (2.18) and (2.20), with probability $1 - N^{-D_1}$, we have

$$|\lambda_{\alpha'} - \gamma_{\alpha'}| \le N^{-2/3+\varepsilon_1}, \qquad \max_{\mu, s_2} |\zeta_\mu(s_2)|^2 \le N^{-1+\varepsilon_1}. \tag{3.3}$$

Hence, throughout the proofs of this section, we always use the scale parameter

$$\eta = N^{-2/3-\varepsilon_0}, \qquad \varepsilon_0 > \varepsilon_1 \text{ is a small constant.} \tag{3.4}$$

3.1. Proof of Lemma 3.1

In a first step, we express the singular vector entries as an integral of Green functions over a random interval, which is recorded as the following lemma.

Lemma 3.2. *Under the assumptions of Lemma 3.1, there exist some small constants $\varepsilon, \delta > 0$ satisfying*

$$\delta > 2\varepsilon, \qquad \varepsilon > C\varepsilon_1, \qquad \delta < C^{-1}\varepsilon_0, \tag{3.5}$$

for some large constant $C > C_1$ (recall (3.2) for C_1) such that

$$\lim_{N\to\infty} \max_{l \le N_k^\delta} \max_{\mu,\nu} \left| \mathbb{E}^V \theta(N\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)) - \mathbb{E}^V \theta\Big(\frac{N}{\pi}\int_I \tilde{G}_{\mu\nu}(z)\mathcal{X}(E)\,\mathrm{d}E\Big)\right| = 0,$$

where I is defined as

$$I := \big[a_{2k-1} - N^{-2/3+\varepsilon},\ a_{2k-1} + N^{-2/3+\varepsilon}\big] \tag{3.6}$$

when (1.11) holds

$$I := \big[a_{2k} - N^{-2/3+\varepsilon}, \qquad a_{2k} + N^{-2/3+\varepsilon}\big]$$

when (1.12) holds. We define

$$\mathcal{X}(E) := \mathbf{1}(\lambda_{\alpha'+1} < E^- \le \lambda_{\alpha'}), \tag{3.7}$$

where $E^{\pm} := E \pm N^\varepsilon \eta$. The conclusion holds if we replace X_V with X_G.

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

Proof. We first observe that

$$\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)=\frac{\eta}{\pi}\int_{\mathbb{R}}\frac{\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)}{(E-\lambda_{\alpha'})^2+\eta^2}\,\mathrm{d}E.$$

Choose a and b such that

$$a:=\min\{\lambda_{\alpha'}-N^{\varepsilon}\eta,\ \lambda_{\alpha'+1}+N^{\varepsilon}\eta\},\qquad b:=\lambda_{\alpha'}+N^{\varepsilon}\eta. \tag{3.8}$$

We also observe the elementary inequality (see the equation above Equation (6.10) of [18]), for some constant $C>0$,

$$\int_x^{\infty}\frac{\eta}{\pi(y^2+\eta^2)}\,\mathrm{d}y\le\frac{C\eta}{x+\eta},\qquad x>0. \tag{3.9}$$

By (3.3), (3.8), and (3.9), with probability $1-N^{-D_1}$, we have

$$\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)=\frac{\eta}{\pi}\int_a^b\frac{\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)}{(E-\lambda_{\alpha}')^2+\eta^2}\,\mathrm{d}E+O(N^{-1-\varepsilon+\varepsilon_1}). \tag{3.10}$$

By (3.2), (3.3), (3.5), (3.10), and mean value theorem, we have

$$\mathbb{E}^V\theta(N\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu))=\mathbb{E}^V\theta\Big(\frac{N\eta}{\pi}\int_a^b\frac{\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)}{(E-\lambda_{\alpha'})^2+\eta^2}\,\mathrm{d}E\Big)+o(1). \tag{3.11}$$

Define $\lambda_t^{\pm}:=\lambda_t\pm N^{\varepsilon}\eta,\ t=\alpha',\ \alpha'+1$, and by (3.8), we have

$$\int_a^b\mathrm{d}E=\int_{\lambda^+_{\alpha'+1}}^{\lambda^+_{\alpha'}}\mathrm{d}E+\mathbf{1}\big(\lambda^+_{\alpha'+1}>\lambda^-_{\alpha'}\big)\int_{\lambda^-_{\alpha'}}^{\lambda^+_{\alpha'+1}}\mathrm{d}E.$$

By (3.2), (3.3), (3.11), and the mean value theorem, we have

$$\mathbb{E}^V\theta(N\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu))=\mathbb{E}^V\theta\Big(\frac{N\eta}{\pi}\int_{\lambda^+_{\alpha'+1}}^{\lambda^+_{\alpha'}}\frac{\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)}{(E-\lambda_{\alpha'})^2+\eta^2}\,\mathrm{d}E\Big)+o(1),$$

where we used (2.18) and (3.5). Next we can, without loss of generality, consider the case when (1.11) holds. By (3.3) and (3.5), we observe that, with probability $1-N^{-D_1}$, we have $\lambda^+_{\alpha'}\le a_{2k-1}+N^{-2/3+\varepsilon}$ and $\lambda^+_{\alpha'+1}\ge a_{2k-1}-N^{-2/3+\varepsilon}$. By (2.18) and the choice of I in (3.6), we have

$$\mathbb{E}^V\theta(N\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu))=\mathbb{E}^V\theta\Big(\frac{N\eta}{\pi}\int_I\frac{\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)}{(E-\lambda_{\alpha'})^2+\eta^2}\mathcal{X}(E)\,\mathrm{d}E\Big)+o(1).$$

Recall (3.1). We can split the summation as

$$\frac{1}{\eta}\tilde{G}_{\mu\nu}(z)=\sum_{\beta\neq\alpha'}\frac{\zeta_{\beta}(\mu)\zeta_{\beta}(\nu)}{(E-\lambda_{\beta})^2+\eta^2}+\frac{\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)}{(E-\lambda_{\alpha'})^2+\eta^2}. \tag{3.12}$$

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

Define $\mathcal{A} := \{\beta \neq \alpha' : \lambda_\beta$ is not in the kth bulk component$\}$. By (3.3), with probability $1 - N^{-D_1}$, we have

$$\Big| \sum_{\beta \neq \alpha'} \frac{N\eta}{\pi} \int_I \frac{\zeta_\beta(\mu)\zeta_\beta(\nu)}{(E-\lambda_\beta)^2+\eta^2} \, \mathrm{d}E \Big|$$

$$\leq \frac{N^{\varepsilon_1}}{\pi} \bigg(\sum_{\beta \in \mathcal{A}} \int_I \frac{\eta}{\eta^2 + (E-\lambda_\beta)^2} \, \mathrm{d}E + \sum_{\beta \in \mathcal{A}^c} \int_I \frac{\eta}{\eta^2 + (E-\lambda_\beta)^2} \, \mathrm{d}E \bigg). \tag{3.13}$$

By Assumption 1.2, with probability $1 - N^{D_1}$, we have

$$\frac{N^{\varepsilon_1}}{\pi} \sum_{\beta \in \mathcal{A}} \int_I \frac{\eta}{\eta^2 + (E-\lambda_\beta)^2} \, \mathrm{d}E \leq N^{\varepsilon_1} \sum_{\beta \in \mathcal{A}} N^{-4/3-\varepsilon_0+\varepsilon}. \tag{3.14}$$

Define

$$l(\beta) := \beta - \sum_{t<k} N_t.$$

By (3.3), with probability $1 - N^{-D_1}$, for some small constant $0 < \delta < 1$, we have

$$\frac{N^{\varepsilon_1}}{\pi} \sum_{\beta \in \mathcal{A}^c} \int_I \frac{\eta}{(E-\lambda_\beta)^2 + \eta^2} \, \mathrm{d}E \leq N^{\varepsilon_1 + \delta} + \frac{1}{\pi} \sum_{\beta \in \mathcal{A}^c,\, l(\beta) \geq N_k^\delta} \int_I \frac{N^{\varepsilon_1}\eta}{\eta^2 + (E-\lambda_\beta)^2} \, \mathrm{d}E. \tag{3.15}$$

By Assumption 1.2, (1.9), (2.18), and the assumption that $\delta > 2\varepsilon$, it is easy to check that (see [24, Equation (3.12)])

$$(E-\lambda_\beta)^2 \geq c \Big(\frac{l(\beta)}{N} \Big)^{4/3}, \qquad c > 0 \text{ is some constant.} \tag{3.16}$$

By (3.16), with probability $1 - N^{-D_1}$, we have

$$\frac{1}{\pi} \sum_{\beta \in \mathcal{A}^c,\, l(\beta) \geq N_k^\delta} \int_I \frac{N^{\varepsilon_1}\eta}{\eta^2 + (E-\lambda_\beta)^2} \, \mathrm{d}E \leq N^{\varepsilon_1 - \varepsilon_0 + \varepsilon} \int_{N^\delta - 1}^{N} \frac{1}{x^{4/3}} \, \mathrm{d}x \leq N^{-\delta/3 + \varepsilon_1 - \varepsilon_0 + \varepsilon}.$$

Recall (3.5). We can restrict $\varepsilon_1 - \varepsilon_0 + \varepsilon < 0$, so that, with probability $1 - N^{-D_1}$, this yields

$$\sum_{\beta \in \mathcal{A}^c,\, l(\beta) \geq N_k^\delta} \int_I \frac{N^{\varepsilon_1}\eta}{\eta^2 + (E-\lambda_\beta)^2} \, \mathrm{d}E \leq N^{-\delta/3}. \tag{3.17}$$

By (3.13), (3.14), (3.15), and (3.17), with probability $1 - N^{-D_1}$, we have

$$\Big| \sum_{\beta \neq \alpha'} \frac{N\eta}{\pi} \int_I \frac{\zeta_\beta(\mu)\zeta_\beta(\nu)}{(E-\lambda_\beta)^2+\eta^2} \, \mathrm{d}E \Big| \leq N^{\delta + 2\varepsilon_1}. \tag{3.18}$$

By (3.2), (3.3), (3.12), (3.18), and the mean value theorem, we have

$$\Big| \mathbb{E}^V \theta \Big(\frac{N\eta}{\pi} \int_I \frac{\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)}{(E-\lambda_{\alpha'})^2+\eta^2} \mathcal{X}(E) \, \mathrm{d}E \Big) - \mathbb{E}^V \theta \Big(\frac{N}{\pi} \int_I \tilde{G}_{\mu\nu}(E+i\eta) \mathcal{X}(E) \, \mathrm{d}E \Big) \Big|$$

$$\leq N^{C_1(\delta + 2\varepsilon_1)} \mathbb{E}^V \sum_{\beta \neq \alpha'} \frac{N\eta}{\pi} \int_I \frac{|\zeta_\beta(\mu)\zeta_\beta(\nu)|}{(E-\lambda_\beta)^2+\eta^2} \mathcal{X}(E) \, \mathrm{d}E, \tag{3.19}$$

where C_1 is defined in (3.2). To complete the proof, it suffices to estimate the right-hand side of (3.19). Similarly to (3.14), we have

$$\sum_{\beta\in\mathcal{A}}\int_I \frac{\eta}{\eta^2+(E-\lambda_\beta)^2}\,\mathrm{d}E \le N^{-1/3-\varepsilon_0+\varepsilon}. \tag{3.20}$$

Choose a small constant $0<\delta_1<1$ and repeat the estimation in (3.17) to obtain

$$\sum_{\beta\in\mathcal{A}^c,\, l(\beta)\ge N_k^{\delta_1}}\int_I \frac{\eta}{\eta^2+(E-\lambda_\beta)^2}\,\mathrm{d}E \le N^{-\delta_1/3+\varepsilon-\varepsilon_0}. \tag{3.21}$$

Recall (1.11), (3.3), and (3.9). Using a discussion similar to that above Equation (3.14) of [24], we conclude that

$$\begin{aligned}\sum_{\beta\in\mathcal{A}^c,\, l\le l(\beta)\le N_k^{\delta_1}} &\frac{N\eta}{\pi}\mathbb{E}^V\int_I \frac{|\zeta_\beta(\mu)\zeta_\beta(\nu)|}{(E-\lambda_\beta)^2+\eta^2}\mathcal{X}(E)\,\mathrm{d}E\\ &\le \mathbb{E}^V\int_{\lambda_{\alpha'+1}+N^{\varepsilon}\eta}^{\infty}\frac{N^{\varepsilon_1}\eta}{(E-\lambda_{\alpha'+1})^2+\eta^2}\,\mathrm{d}E\\ &\le N^{-\varepsilon+\varepsilon_1},\end{aligned} \tag{3.22}$$

where we have used the fact that $\beta\in\mathcal{A}^c$ and $l<l(\beta)\le N_k^{\delta_1}$ imply that $\lambda_\beta\le\lambda_{\alpha'+1}$. It is notable that the above bound is independent of δ. It remains to estimate the summation of the terms when $\beta\in\mathcal{A}^c$ and $l(\beta)<l$. For a given constant, ε' satisfies

$$\delta>2\varepsilon', \qquad \varepsilon'>C\varepsilon_1, \qquad \delta<C^{-1}\varepsilon_0. \tag{3.23}$$

We partition $I=I_1\cup I_2$ with $I_1\cap I_2=\varnothing$, where

$$I_1 := \{E\in I:\ \text{there exists } \beta,\ \beta\in\mathcal{A}^c,\ \ l(\beta)<l,\ \ |E-\lambda_\beta|\le N^{\varepsilon'}\eta\}. \tag{3.24}$$

By (3.3) and (3.24), using a similar discussion to that used for (3.22), we have

$$\sum_{\beta\in\mathcal{A}^c;\ l(\beta)<l}\frac{N\eta}{\pi}\mathbb{E}^V\int_{I_2}\frac{|\zeta_\beta(\mu)\zeta_\beta(\nu)|}{(E-\lambda_\beta)^2+\eta^2}\mathcal{X}(E)\,\mathrm{d}E\le N^{-2\varepsilon'+\varepsilon_1}.$$

It is easy to check that on I_1 when $\lambda_{\alpha'+1}\le\lambda_{\alpha'}<\lambda_\beta$, we have (see (3.15) of [24])

$$\frac{1}{(E-\lambda_\beta)^2+\eta^2}\mathbf{1}(E^-\le\lambda_{\alpha'})\le\frac{N^{2\varepsilon}}{(\lambda_{\alpha'+1}-\lambda_{\alpha'})^2+\eta^2}. \tag{3.25}$$

By Lemma 2.2, the above equation holds with probability $1-N^{-D_1}$. By (3.3), (3.25), and a discussion similar to that used in [24, Equation (3.16)], we have

$$\begin{aligned}\sum_{\beta\in\mathcal{A}^c,\, l(\beta)\le l}\frac{N\eta}{\pi}\mathbb{E}^V\int_{I_1}\frac{|\zeta_\beta(\mu)\zeta_\beta(\nu)|}{(E-\lambda_\beta)^2+\eta^2}\mathcal{X}(E)\,\mathrm{d}E &\le \mathbb{E}^V\int_{I_1}\frac{N^{\varepsilon_1+2\varepsilon}\eta^2}{(\lambda_{\alpha'+1}-\lambda_{\alpha'})^2+\eta^2}\,\mathrm{d}E\\ &\le \mathbb{E}^V\mathbf{1}(|\lambda_{\alpha'+1}-\lambda_{\alpha'}|\le N^{-1/3}\eta^{1/2})+N^{-D_1+\varepsilon_1+3\varepsilon}\\ &\le N^{-\varepsilon_0/2+3\varepsilon}.\end{aligned} \tag{3.26}$$

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

By (3.20), (3.21), (3.22), (3.23), and (3.26), we conclude the proof of (3.19). It is clear that our proof still applies when we replace X_V with X_G. □

In a second step, we write the sharp indicator function of (3.7) as a some smooth function q of $\tilde{G}_{\mu\nu}$. To be consistent with the proof of Lemma 3.2, we consider the bulk edge a_{2k-1}. Define

$$\vartheta_\eta(x) := \frac{\eta}{\pi(x^2+\eta^2)} = \frac{1}{\pi}\operatorname{Im}\frac{1}{x-\mathrm{i}\eta}.$$

We define a smooth cutoff function $q \equiv q_{\alpha'} : \mathbb{R} \to \mathbb{R}_+$ as

$$q(x) = \begin{cases} 1 & \text{if } |x-l| \le \frac{1}{3}, \\ 0 & \text{if } |x-l| \ge \frac{2}{3}, \end{cases} \tag{3.27}$$

where l is defined in (1.11). We also let $Q_1 = Y^*Y$.

Lemma 3.3. *For ε given in (3.5), define*

$$\mathcal{X}_E(x) := \mathbf{1}(E^- \le x \le E_U), \tag{3.28}$$

where $E_U := a_{2k-1} + 2N^{-2/3+\varepsilon}$, and define $\tilde{\eta} := N^{-2/3-9\varepsilon_0}$, where ε_0 is defined in (3.4). Then

$$\lim_{N\to\infty}\max_{l\le N_k^\delta}\max_{\mu,\nu}\left|\mathbb{E}^V\theta\Big(N\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)\Big) - \mathbb{E}^V\theta\Big(\frac{N}{\pi}\int_I \tilde{G}_{\mu\nu}(z)q\big[\operatorname{Tr}(\mathcal{X}_E * \vartheta_{\tilde{\eta}})(Q_1)\big]\,\mathrm{d}E\Big)\right| = 0,$$

where I is defined in (3.6) and '$*$' is the convolution operator.

Proof. For any $E_1 < E_2$, denote the number of eigenvalues of Q_1 in $[E_1, E_2]$ by

$$\mathcal{N}(E_1, E_2) := \#\big\{j : E_1 \le \lambda_j \le E_2\big\}. \tag{3.29}$$

Recall (3.6) and (3.7). It is easy to check that, with probability $1 - N^{-D_1}$, we have

$$\begin{aligned} N\int_I \tilde{G}_{\mu\nu}(z)\mathcal{X}(E)\,\mathrm{d}E &= N\int_I \tilde{G}_{\mu\nu}(z)\mathbf{1}(\mathcal{N}(E^-, E_U) = l)\,\mathrm{d}E \\ &= N\int_I \tilde{G}_{\mu\nu}(z)q[\operatorname{Tr}\mathcal{X}_E(Q_1)]\,\mathrm{d}E, \end{aligned} \tag{3.30}$$

where, for the second equality, we used (2.18) and Assumption 1.2. We use the following lemma to estimate (3.29) by its delta approximation smoothed on the scale $\tilde{\eta}$. The proof is given in the supplementary material [14].

Lemma 3.4. *For $t = N^{-2/3-3\varepsilon_0}$, there exists some constant C, and with probability $1 - N^{-D_1}$, for any E satisfying*

$$|E^- - a_{2k-1}| \le \frac{3}{2}N^{-2/3+\varepsilon},$$

we have

$$|\operatorname{Tr}\mathcal{X}_E(Q_1) - \operatorname{Tr}(\mathcal{X}_E * \vartheta_{\tilde{\eta}})(Q_1)| \le C\big(N^{-2\varepsilon_0} + \mathcal{N}(E^- - t,\ E^- + t)\big). \tag{3.31}$$

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

By Equation (A.7) of [26], for any $z \in \boldsymbol{D}(\tau)$ defined in (2.1), we have

$$\operatorname{Im} m(z) \sim \begin{cases} \dfrac{\eta}{\sqrt{\kappa+\eta}}, & E \notin \operatorname{supp}(\rho), \\ \sqrt{\kappa+\eta}, & E \in \operatorname{supp}(\rho), \end{cases} \tag{3.32}$$

where $\kappa := |E - a_{2k-1}|$. When $\mu = \nu$, with probability $1 - N^{-D_1}$, we have

$$\begin{aligned} \sup_{E \in I} |\tilde{G}_{\mu\mu}(E + \mathrm{i}\eta)| &= \sup_{E \in I} |\operatorname{Im} G_{\mu\mu}(z)| \\ &\leq \sup_{E \in I} \big(\operatorname{Im} |G_{\mu\mu}(z) - m(z)| + |\operatorname{Im} m(z)| \big) \\ &\leq N^{-1/3+\varepsilon_0+2\varepsilon}, \end{aligned}$$

where we have used (2.17) and (3.32). When $\mu \neq \nu$, we use the identity

$$\tilde{G}_{\mu\nu} = \eta \sum_{k=M+1}^{M+N} G_{\mu k} \overline{G}_{\nu k}.$$

By (2.17) and (3.32), with probability $1 - N^{-D_1}$, we have $\sup_{E \in I} |\tilde{G}_{\mu\nu}(z)| \leq N^{-1/3+\varepsilon_0+2\varepsilon}$. Therefore, for $E \in I$, with probability $1 - N^{-D_1}$, we have

$$\sup_{E \in I} |\tilde{G}_{\mu\nu}(E + \mathrm{i}\eta)| \leq N^{-1/3+3\varepsilon_0/2}. \tag{3.33}$$

Recall (3.27). By (3.30), (3.31), (3.33), and the smoothness of q, with probability $1 - N^{-D_1}$, we have

$$\begin{aligned} &\left| N \int_I \tilde{G}_{\mu\nu}(z) \mathcal{X}(E)\, \mathrm{d}E - N \int_I \tilde{G}_{\mu\nu}(z) q[\operatorname{Tr}(\mathcal{X}_E * \vartheta_{\tilde{\eta}}(Q_1))]\, \mathrm{d}E \right| \\ &\quad \leq CN \sum_{l(\beta) \leq N_k^{\delta}} \int_I |\tilde{G}_{\mu\nu}(z)| \mathbf{1}(|E^- - \lambda_\beta| \leq t)\, \mathrm{d}E + N^{-\varepsilon_0/4} \\ &\quad \leq CN^{1+\delta} |t| \sup_{z \in I} |\tilde{G}_{\mu\nu}(z)| + N^{-\varepsilon_0/4}. \end{aligned} \tag{3.34}$$

By (3.33) and (3.34), we have

$$\left| N \int_I \tilde{G}_{\mu\nu}(z) \mathcal{X}(E)\, \mathrm{d}E - N \int_I \tilde{G}_{\mu\nu}(z) q\big[\operatorname{Tr}(\mathcal{X}_E * \vartheta_{\tilde{\eta}}(Q_1))\big]\, \mathrm{d}E \right| \leq CN^{-\varepsilon_0/2+\delta} + N^{-\varepsilon_0/4}.$$

Using a discussion similar to that used for (3.13), by (3.2) and (3.5), we complete the proof. □

In the final step, we use the Green function comparison argument to prove the following lemma, whose proof is given in Section 3.2.

Lemma 3.5. *Under the assumptions of Lemma 3.3, we have*

$$\lim_{N \to \infty} \max_{\mu,\nu} (\mathbb{E}^V - \mathbb{E}^G) \theta \Big(\frac{N}{\pi} \int_I \tilde{G}_{\mu\nu}(z) q\big[\operatorname{Tr}(\mathcal{X}_E * \vartheta_{\tilde{\eta}})(Q_1)\big]\, \mathrm{d}E \Big) = 0.$$

The proof of Lemma 3.1 follows from the proof of Lemma 3.3.

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

3.2. The Green function comparsion argument

In this section we prove Lemma 3.5 using the Green function comparison argument. At the end of this section we discuss how we can extend Lemma 3.1 to Theorem 1.1 and Theorem 1.2. By the orthonormal properties of ξ and ζ, and (2.6), we have

$$\tilde{G}_{ij} = \eta \sum_{k=1}^{M} G_{ik}\overline{G}_{jk}, \qquad \tilde{G}_{\mu\nu} = \eta \sum_{k=M+1}^{M+N} G_{\mu k}\overline{G}_{\nu k}. \tag{3.35}$$

By (2.17), with probability $1 - N^{-D_1}$, we have

$$|G_{\mu\mu}| = O(1), \ |G_{\mu\nu}| \leq N^{-1/3+2\varepsilon_0}, \qquad \mu \neq \nu. \tag{3.36}$$

We first drop the all diagonal terms in (3.35).

Lemma 3.6. *Recall that $E_U = a_{2k-1} + 2N^{-2/3+\varepsilon}$ and $\tilde{\eta} = N^{-2/3-9\varepsilon_0}$. We have*

$$\mathbb{E}^V \theta\Big[\frac{N}{\pi}\int_I \tilde{G}_{\mu\nu}(z) q\big[\mathrm{Tr}\,(\mathcal{X}_E * \vartheta_{\tilde{\eta}})(Q_1)\big]\,\mathrm{d}E\Big] - \mathbb{E}^V \theta\Big[\int_I x(E) q(y(E))\,\mathrm{d}E\Big] = o(1), \tag{3.37}$$

where

$$x(E) := \frac{N\eta}{\pi} \sum_{k=M+1,\, k\neq\mu,\nu}^{M+N} X_{\mu\nu,k}(E + \mathrm{i}\eta), \qquad y(E) := \frac{\tilde{\eta}}{\pi}\int_{E^-}^{E_U} \sum_k \sum_{\beta\neq k} X_{\beta\beta,k}(E+\mathrm{i}\tilde{\eta})\,\mathrm{d}E, \tag{3.38}$$

and $X_{\mu\nu,k} := G_{\mu k}\overline{G}_{\nu k}$. The conclusion holds if we replace X_V with X_G.

Proof. We first observe that, by (3.36), with probability $1 - N^{-D_1}$, we have

$$|x(E)| \leq N^{2/3+3\varepsilon_0}, \tag{3.39}$$

which implies that

$$\int_I |x(E)|\,\mathrm{d}E \leq N^{4\varepsilon_0}. \tag{3.40}$$

By (3.35) and (3.36), with probability $1 - N^{-D_1}$, we have

$$\begin{aligned}\Big|\frac{N}{\pi}\tilde{G}_{\mu\nu}(E+\mathrm{i}\eta) - x(E)\Big| &= \frac{N\eta}{\pi}|G_{\mu\mu}\overline{G}_{\nu\mu} + G_{\mu\nu}\overline{G}_{\nu\nu}| \\ &\leq N\eta\big(\mathbf{1}(\mu=\nu) + N^{-1/3+2\varepsilon_0}\mathbf{1}(\mu\neq\nu)\big).\end{aligned} \tag{3.41}$$

By Equations (5.11) and (6.42) of [16], we have

$$\mathrm{Tr}\,(\mathcal{X}_E * \vartheta_{\tilde{\eta}}(Q_1)) = \frac{N}{\pi}\int_{E^-}^{E_U} \mathrm{Im}\, m_2(w + \mathrm{i}\tilde{\eta})\,\mathrm{d}w, \qquad \sum_{\mu\nu} |G_{\mu\nu}(w+\mathrm{i}\tilde{\eta})|^2 = \frac{N\,\mathrm{Im}\, m_2(w+\mathrm{i}\tilde{\eta})}{\tilde{\eta}}. \tag{3.42}$$

Therefore, we have

$$\mathrm{Tr}\,(\mathcal{X}_E * \vartheta_{\tilde{\eta}}(Q_1)) - y(E) = \frac{\tilde{\eta}}{\pi}\int_{E^-}^{E_U} \sum_{\beta=M+1}^{M+N} |G_{\beta\beta}|^2\,\mathrm{d}w. \tag{3.43}$$

By (3.43), the mean value theorem, and the fact that q is smooth enough, we have

$$\left|q\left[\operatorname{Tr}(\mathcal{X}_E * \vartheta_{\tilde{\eta}})(Q_1)\right] - q[y(E)]\right| \le N^{-1/3-7\varepsilon_0}. \tag{3.44}$$

Therefore, by the mean value theorem, (3.2), (3.5), (3.39), (3.40), (3.41), and (3.44), we complete the proof. □

To prove Lemma 3.5, by (3.37), it suffices to prove that

$$[\mathbb{E}^V - \mathbb{E}^G]\theta\left(\int_I x(E)q(y(E))\,\mathrm{d}E\right) = o(1). \tag{3.45}$$

We use the Green function comparison argument to prove (3.45), where we follow the basic approach of [16, Section 6] and [24, Section 3.1]. Define a bijective ordering map Φ on the index set, where

$$\Phi\colon \left\{(i, \mu_1)\colon 1 \le i \le M,\ M+1 \le \mu_1 \le M+N\right\} \to \{1, \dots, \gamma_{\max} = MN\}.$$

Recall that we relabel $X^V = ((X_V)_{i\mu_1}, i \in \mathcal{I}_1,\ \mu_1 \in \mathcal{I}_2)$, and similarly for X^G. For any $1 \le \gamma \le \gamma_{\max}$, we define the matrix $X_\gamma = \left(x^\gamma_{i\mu_1}\right)$ such that $x^\gamma_{i\mu_1} = X^G_{i\mu_1}$ if $\Phi(i, \mu_1) > \gamma$ and $x^\gamma_{i\mu_1} = X^V_{i\mu_1}$ otherwise. Note that $X_0 = X^G$ and $X_{\gamma_{\max}} = X^V$. With the above definitions, we have

$$\left[\mathbb{E}^G - \mathbb{E}^V\right]\theta\left(\int_I x(E)q(y(E))\,\mathrm{d}E\right) = \sum_{\gamma=1}^{\gamma_{\max}} [\mathbb{E}^{\gamma-1} - \mathbb{E}^\gamma]\theta\left(\int_I x(E)q(y(E))\,\mathrm{d}E\right).$$

For simplicity, we rewrite the above equation as

$$\begin{aligned}\mathbb{E}\left[\theta\left(\int_I x^G q(y^G)\,\mathrm{d}E\right) - \theta\left(\int_I x^V q(y^V)\,\mathrm{d}E\right)\right]\\ = \sum_{\gamma=1}^{\gamma_{\max}} \mathbb{E}\left[\theta\left(\int_I x_{\gamma-1}q(y_{\gamma-1})\,\mathrm{d}E\right) - \theta\left(\int_I x_\gamma q(y_\gamma)\,\mathrm{d}E\right)\right].\end{aligned}$$

The key step of the Green function comparison argument is to use the Lindeberg replacement strategy. We focus on the indices $s, t \in \mathcal{I}$; the special case $\mu, \nu \in \mathcal{I}_2$ follows. Define $Y_\gamma := \Sigma^{1/2}X_\gamma$ and

$$H^\gamma := \begin{pmatrix} 0 & z^{1/2}Y_\gamma \\ z^{1/2}Y_\gamma^* & 0 \end{pmatrix}, \qquad G^\gamma := \begin{pmatrix} -zI & z^{1/2}Y_\gamma \\ z^{1/2}Y_\gamma^* & -zI \end{pmatrix}^{-1}. \tag{3.46}$$

As Σ is diagonal, for each fixed γ, H^γ and $H^{\gamma-1}$ differ only at the (i, μ_1) and (μ_1, i) elements, where $\Phi(i, \mu_1) = \gamma$. Then we define the $(N+M) \times (N+M)$ matrices V and W by

$$\begin{aligned} V_{ab} &= z^{1/2}\left(\mathbf{1}_{\{(a,b)=(i,\mu_1)\}} + \mathbf{1}_{\{(a,b)=(\mu_1,i)\}}\right)\sqrt{\sigma_i}X^G_{i\mu_1},\\ W_{ab} &= z^{1/2}\left(\mathbf{1}_{\{(a,b)=(i,\mu_1)\}} + \mathbf{1}_{\{(a,b)=(\mu_1,i)\}}\right)\sqrt{\sigma_i}X^V_{i\mu_1},\end{aligned}$$

so that H^γ and $H^{\gamma-1}$ can be written as

$$H^{\gamma-1} = O + V, \qquad H^\gamma = O + W,$$

for some $(N+M)\times(N+M)$ matrix O satisfying $O_{i\mu_1}=O_{\mu_1 i}=0$, with O independent of V and W. Define

$$S:=(H^{\gamma-1}-z)^{-1}, \qquad R:=(O-z)^{-1}, \qquad T:=(H^{\gamma}-z)^{-1}. \tag{3.47}$$

With the above definitions, we can write

$$\begin{aligned}\mathbb{E}\Big[\theta\Big(\int_I x^G q(y^G)\,\mathrm{d}E\Big)-\theta\Big(\int_I x^V q(y^V)\,\mathrm{d}E\Big)\Big] \\ =\sum_{\gamma=1}^{\gamma_{\max}}\mathbb{E}\Big[\theta\Big(\int_I x^S q(y^S)\,\mathrm{d}E\Big)-\theta\Big(\int_I x^T q(y^T)\,\mathrm{d}E\Big)\Big].\end{aligned} \tag{3.48}$$

The comparison argument is based on the resolvent expansion

$$S=R-RVR+(RV)^2R-(RV)^3R+(RV)^4S. \tag{3.49}$$

For any integer $m>0$, by Equation (6.11) of [16], we have

$$([RV]^m R)_{ab}=\sum_{(a_i,b_i)\in\{(i,\mu_1),(\mu_1,i)\},\ 1\le i\le m}(z)^{m/2}(\sigma_i)^{m/2}\big(X^G_{i\mu_1}\big)^m R_{aa_1}R_{b_1a_2}\ldots R_{b_mb}, \tag{3.50}$$

$$([RV]^m S)_{ab}=\sum_{(a_i,b_i)\in\{(i,\mu_1),(\mu_1,i)\},\ 1\le i\le m}(z)^{m/2}(\sigma_i)^{m/2}\big(X^G_{i\mu_1}\big)^m R_{aa_1}R_{b_1a_2}\ldots S_{b_mb}. \tag{3.51}$$

Define

$$\Delta X_{\mu\nu,k}:=S_{\mu k}\overline{S}_{\nu k}-R_{\mu k}\overline{R}_{\nu k}. \tag{3.52}$$

In [24], the discussion relied on a crucial parameter (see [24, Equation (3.32)]), which counts the maximum number of diagonal resolvent elements in $\Delta X_{\mu\nu,k}$. We will follow this strategy using a different counting parameter, and, furthermore, use (3.50) and (3.51) as our key ingredients. Our discussion is slightly easier due to the loss of a free index (i.e. $i\neq\mu_1$).

Inserting (3.49) into (3.52), by (3.50) and (3.51), we find that there exists a random variable A_1, which depends on the randomness only through O and the first two moments of $X^G_{i\mu_1}$. Taking the partial expectation with respect to the $(i,\,\mu_1)$th entry of X^G(recall they are i.i.d.), by (1.2), we have the following result.

Lemma 3.7. *Recall (2.7), and let $\mathbb{E}_\gamma$ be the partial expectation with respect to $X^G_{i\mu_1}$. Then there exists some constant $C>0$, and with probability $1-N^{-D_1}$, we have*

$$\big|\mathbb{E}_\gamma\,\Delta X_{\mu\nu,k}-A_1\big|\le N^{-3/2+C\varepsilon_0}\Psi(z)^{3-s}, \qquad M+1\le k\neq\mu,\,\nu\le M+N,$$

where s counts the maximum number of resolvent elements in $\Delta X_{\mu\nu,k}$ involving the index μ_1 and is defined as

$$s:=\mathbf{1}\big((\{\mu,\,\nu\}\cap\{\mu_1\}\neq\varnothing)\cup(\{k=\mu_1\})\big). \tag{3.53}$$

Proof. Inserting (3.49) into (3.52), the terms in the expansion containing $X^G_{i\mu_1}$ and $(X^G_{i\mu_1})^2$ will be included in A_1; we consider only the terms containing $(X^G_{i\mu_1})^m$, $m\ge 3$. We consider $m=3$ and discuss the terms

$$R_{\mu k}\overline{[(RV)^3R]}_{\nu k}, \qquad [RVR]_{\mu k}\overline{[(RV)^2R]}_{\nu k}.$$

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

By (3.50), we have

$$R_{\mu k}\overline{[(RV)^3R]_{\nu k}} = R_{\mu k}\Big(\sum (\sigma_i)^{3/2}(X^G_{i\mu_1})^3\overline{(z)^{3/2}R_{\nu a_1}R_{b_1a_2}R_{b_2a_3}R_{b_3k}}\Big).$$

In the worst scenario, $R_{b_1a_2}$ and $R_{b_2a_3}$ are assumed to be the diagonal entries of R. Similarly, we have

$$[RVR]_{\mu k}\overline{[(RV)^2R]_{\nu k}} = \Big(\sum z^{1/2}\sigma_i^{1/2}X^G_{i\mu_1}R_{\mu a_1}R_{b_1k}\Big)\Big(\sum \sigma_i(X^G_{i\mu_1})^2\overline{zR_{\nu a_1}R_{b_1a_2}R_{b_2k}}\Big),$$

and the worst scenario is the case when $R_{b_1a_2}$ is a diagonal term. As $\mu, \nu \neq i$ always holds and there are only a finite number of terms in the summation, by (1.2) and (3.36), for some constant C, we have

$$\mathbb{E}_\gamma|R_{\mu k}\overline{[(RV)^3R]_{\nu k}}| \leq N^{-3/2+C\varepsilon_0}\Psi(z)^{3-s}.$$

Similarly, we have

$$\mathbb{E}_\gamma|[RVR]_{\mu k}\overline{[(RV)^2R]_{\nu k}}| \leq N^{-3/2+C\varepsilon_0}\Psi(z)^{3-s}.$$

The cases in which $4 \leq m \leq 8$ can be handled similarly. This completes the proof. □

Lemma 3.5 follows from the following lemma. Recall (3.38), and define

$$\Delta x(E) := x^S(E) - x^R(E), \qquad \Delta y(E) := y^S(E) - y^R(E).$$

Lemma 3.8. *For any fixed μ, ν, and γ, there exists a random variable A, which depends on the randomness only through O and the first two moments of X^G, such that*

$$\mathbb{E}\theta\Big(\int_I x^S q(y^S)\,\mathrm{d}E\Big) - \mathbb{E}\theta\Big(\int_I x^R q(y^R)\,\mathrm{d}E\Big) = A + o(N^{-2+t}), \tag{3.54}$$

where $t := |\mu, \nu \cap \mu_1|$.

The proof of Lemma 3.8 given in the supplementary material [14]. We now show how Lemma 3.8 implies Lemma 3.5.

Proof of Lemma 3.5. It is easy to check that Lemma 3.8 still holds when we replace S with T. Note that in (3.48) there are $O(N)$ terms when $t = 1$ and $O(N^2)$ terms when $t = 0$. By (3.54), we have

$$\mathbb{E}\Big[\theta\Big(\int_I x^G q(y^G)\,\mathrm{d}E\Big) - \theta\Big(\int_I x^V q(y^V)\,\mathrm{d}E\Big)\Big] = o(1),$$

where we have used the assumption that the first two moments of X^V are the same as those of X^G. Combine with (3.37) to complete the proof. □

It is clear that our proof can be extended to the left singular vectors. For the proof of Theorem 1.1, the only difference is that we use the mean value theorem in $\mathbb{R}^2$ whenever it is needed. Moreover, for the proof of Theorem 1.2, we need to use n intervals defined by

$$I_i := \big[a_{2k_i-1} - N^{-2/3+\varepsilon},\, a_{2k_i-1} + N^{-2/3+\varepsilon}\big], \qquad i = 1, 2, \ldots, n.$$

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

3.3. Extension to singular values

In this section we discuss how the arguments of Section 3.2 can be applied to the general function θ defined in (1.15) containing singular values. We mainly focus on discussing the proof of Corollary 1.1.

On the one hand, similarly to Lemma 3.3, we can write the singular values in terms of an integral of smooth functions of Green functions. Using the comparison argument with $\theta \in \mathbb{R}^3$ and the mean value theorem in $\mathbb{R}^3$ completes our proof. Similar discussions and results have been derived in [18, Corollary 6.2 and Theorem 6.3]. For completeness, we basically follow the strategy in [24, Section 4] to prove Corollary 1.1. The basic idea is to write the function θ in terms of Green functions by using integration by parts. We mainly look at the right edge of the kth bulk component.

Proof of Corollary 1.1. Let F^V be the law of $\lambda_{\alpha'}$, and consider a smooth function $\theta : \mathbb{R} \to \mathbb{R}$. For δ defined in Lemma 3.2, when $l \leq N_k^{\delta}$, by (1.14) and (2.18), it is easy to check that

$$\mathbb{E}^V \theta\Big(\frac{N^{2/3}}{\varpi}(\lambda_{\alpha'} - a_{2k-1})\Big) = \int_I \theta\Big(\frac{N^{2/3}}{\varpi}(E - a_{2k-1})\Big)\, \mathrm{d}F^V(E) + O(N^{-D_1}), \tag{3.55}$$

where $\varpi := \varpi_{2k-1}$ and I is defined in (3.6). Using integration by parts on (3.55), we have

$$\begin{aligned}[][\mathbb{E}^V - \mathbb{E}^G]\theta\Big(\frac{N^{2/3}}{\varpi}(\lambda_{\alpha'} - a_{2k-1})\Big) \\ = -[\mathbb{E}^V - \mathbb{E}^G]\int_I \frac{N^{2/3}}{\varpi}\theta'\Big(\frac{N^{2/3}}{\varpi}(E - a_{2k-1})\Big)\mathbf{1}(\lambda_{\alpha'} \leq E)dE + O(N^{-D_1}),\end{aligned} \tag{3.56}$$

where we have used (1.14) and (2.18). Similarly to (3.27), recalling (1.11), choose a smooth nonincreasing function f_l that vanishes on the interval $[l + \frac{2}{3}, \infty)$ and is equal to 1 on the interval $(-\infty, l + \frac{1}{3}]$. Recall that $E_U = a_{2k-1} + 2N^{-2/3+\varepsilon}$ and $\mathcal{N}(E, E_U)$ denotes the number of eigenvalues of Q_1 located in the interval $[E, E_U]$. By (3.56), we have

$$\begin{aligned}[][\mathbb{E}^V - \mathbb{E}^G]\theta\Big(\frac{N^{2/3}}{\varpi}(\lambda_{\alpha'} - a_{2k-1})\Big) \\ = -[\mathbb{E}^V - \mathbb{E}^G]\int_I \frac{N^{2/3}}{\varpi}\theta'\Big(\frac{N^{2/3}}{\varpi}(E - a_{2k-1})\Big)f_l(\mathcal{N}(E, E_U))dE + O(N^{-D_1}).\end{aligned}$$

Recall that $\tilde{\eta} = N^{-2/3-9\varepsilon_0}$. Similarly to the discussion of (3.31), with probability $1 - N^{-D_1}$, we have

$$N^{2/3}\int_I \Big| \operatorname{Tr}\big(\mathbf{1}_{[E,E_U]} * \vartheta_{\tilde{\eta}}(Q_1)\big) - \operatorname{Tr}\big(\mathbf{1}_{[E,E_U]}(Q_1)\big)\Big|\, \mathrm{d}E \leq N^{-\varepsilon_0}.$$

This yields

$$\begin{aligned}[][\mathbb{E}^V - \mathbb{E}^G]\theta\Big(\frac{N^{2/3}}{\varpi}(\lambda_{\alpha'} - a_{2k-1})\Big) \\ = -[\mathbb{E}^V - \mathbb{E}^G]\int_I \frac{N^{2/3}}{\varpi}\theta'\Big(\frac{N^{2/3}}{\varpi}(E - a_{2k-1})\Big)f_l\Big(\operatorname{Tr}\big(\mathbf{1}_{[E,E_U]} * \vartheta_{\tilde{\eta}}(Q_1)\big)\Big)\, \mathrm{d}E + O(N^{-D_1}).\end{aligned}$$

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

Integration by parts yields

$$[\mathbb{E}^V - \mathbb{E}^G]\theta\Big(\frac{N^{2/3}}{\varpi}(\lambda_{\alpha'} - a_{2k-1})\Big)$$
$$= \frac{N}{\pi}[\mathbb{E}^V - \mathbb{E}^G]\int_I \theta\Big(\frac{N^{2/3}}{\varpi}(\lambda_{\alpha'} - a_{2k-1})\Big)$$
$$\times f_l'\big(\operatorname{Tr}(\mathbf{1}_{[E,E_U]} * \vartheta_{\tilde{\eta}}(Q_1))\big)\operatorname{Im} m_2(E + \mathrm{i}\tilde{\eta})\,\mathrm{d}E + o(1),$$

where we have used (3.42). Now we extend θ to the general case defined in (1.15). By Theorem 1.1, it is easy to check that

$$[\mathbb{E}^V - \mathbb{E}^G]\theta\Big(\frac{N^{2/3}}{\varpi}(\lambda_{\alpha'} - a_{2k-1}), N\xi_{\alpha'}(i)\xi_{\alpha'}(j), N\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)\Big)$$
$$= \frac{1}{\pi}[\mathbb{E}^V - \mathbb{E}^G]\int_I \theta\Big(\frac{N^{2/3}}{\varpi}(\lambda_{\alpha'} - a_{2k-1}), \phi_{\alpha'}, \varphi_{\alpha'}\Big)$$
$$\times f_l'\big(\operatorname{Tr}(\mathbf{1}_{[E,E_U]} * \vartheta_{\tilde{\eta}}(Q_1))\big) N \operatorname{Im} m_2(E + \mathrm{i}\tilde{\eta})\,\mathrm{d}E + o(1), \tag{3.57}$$

where

$$\phi_{\alpha'} = \frac{N}{\pi}\int_I \tilde{G}_{ij}(\tilde{E} + i\eta) q_1\big[\operatorname{Tr}(\mathbf{1}_{[\tilde{E}^-, E_U]} * \vartheta_{\tilde{\eta}}(Q_1))\big]\,\mathrm{d}\tilde{E},$$
$$\varphi_{\alpha'} = \frac{N}{\pi}\int_I \tilde{G}_{\mu\nu}(\tilde{E} + i\eta) q_2\big[\operatorname{Tr}(\mathbf{1}_{[\tilde{E}^-, E_U]} * \vartheta_{\tilde{\eta}}(Q_1))\big] d\tilde{E},$$

and q_1 and q_2 are the functions defined in (3.27). Therefore, the randomness on the right-hand side of (3.57) is expressed in terms of Green functions. Hence, we can apply the Green function comparison argument to (3.57) as in Section 3.2. The complications are notational and we will not reproduce the details here. □

Finally, the proof of Corollary 1.2 is very similar to that of Corollary 1.1 except that we use n different intervals and a multidimensional integral. We will not reproduce the details here.

4. Singular vectors in the bulks

In this section we prove the bulk universality Theorems 1.3 and 1.4. Our key ingredients, Lemmas 2.1 and 2.4 and Corollary 2.1, are proved for $N^{-1+\tau} \le \eta \le \tau^{-1}$ (recall (2.1)). In the bulks, recalling Lemma 2.3, the eigenvalue spacing is of order N^{-1}. The following lemma extends the above controls for a small spectral scale all the way down to the real axis. The proof relies on Corollary 2.1 and the details can be found in [24, Lemma 5.1].

Lemma 4.1. *Recall (2.19). For $z \in \boldsymbol{D}_k^b$ with $0 < \eta \le \tau^{-1}$, when N is large enough, with probability $1 - N^{-D_1}$, we have*

$$\max_{\mu,\nu} |G_{\mu\nu} - \delta_{\mu\nu} m(z)| \le N^{\varepsilon_1}\Psi(z). \tag{4.1}$$

Once Lemma 4.1 is established, Lemmas 2.3 and 2.4 will follow. Next we follow the basic proof strategy for Theorem 1.1, but use a different spectral window size. Again, we provide only the proof of Lemma 4.2 below, which establishes the universality for the distribution of $\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)$ in detail. Throughout this section, we use the scale parameter

$$\eta = N^{-1-\varepsilon_0}, \qquad \varepsilon_0 > \varepsilon_1 \text{ is a small constant.} \tag{4.2}$$

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

Therefore, the following bounds hold with probability $1 - N^{-D_1}$.

$$\max_{\mu} |G_{\mu\mu}(z)| \le N^{2\varepsilon_0}, \qquad \max_{\mu \neq \nu} |G_{\mu\nu}(z)| \le N^{2\varepsilon_0}, \qquad \max_{\mu,s} |\zeta_\mu(s)|^2 \le N^{-1+\varepsilon_0}. \tag{4.3}$$

The following lemma states the bulk universality for $\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)$.

Lemma 4.2. *Suppose that $Q_V = \Sigma^{1/2} X_V X_V^* \Sigma^{1/2}$ satisfies Assumption 1.1. Assume that the third and fourth moments of X_V agree with those of X_G, and consider the kth, $k = 1, 2, \ldots, p$ bulk component, with l defined in (1.11) or (1.12). Under Assumptions 1.2 and 1.3, for any choices of indices $\mu, \nu \in \mathcal{I}_2$, there exists a small $\delta \in (0, 1)$ such that, when $\delta N_k \le l \le (1-\delta)N_k$, we have*

$$\lim_{N\to\infty} [\mathbb{E}^V - \mathbb{E}^G]\theta(N\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)) = 0,$$

where θ is a smooth function in $\mathbb{R}$ that satisfies

$$|\theta^{(5)}(x)| \le C_1(1+|x|)^{C_1} \quad \text{with some constant } C_1 > 0. \tag{4.4}$$

4.1. Proof of Lemma 4.2

The proof strategy is very similar to that of Lemma 3.1. Our first step is an analogue of Lemma 3.2. The proof is quite similar (actually easier as the window size is much smaller). We omit further details.

Lemma 4.3. *Under the assumptions of Lemma 4.2, there exists a $0 < \delta < 1$ such that*

$$\lim_{N\to\infty} \max_{\delta N_k \le l \le (1-\delta)N_k} \max_{\mu,\nu} \left| \mathbb{E}^V \theta(N\zeta_{\alpha'}(\mu)\zeta_{\alpha'}(\nu)) - \mathbb{E}^V \theta\left[\frac{N}{\pi}\int_I \tilde{G}_{\mu\nu}(z)\mathcal{X}(E)\,\mathrm{d}E\right]\right| = 0, \tag{4.5}$$

where $\mathcal{X}(E)$ is defined in (3.7) and, for ε satisfying (3.5),

$$I := \left[\gamma_{\alpha'} - N^{-1+\varepsilon},\ \gamma_{\alpha'} + N^{-1+\varepsilon}\right]. \tag{4.6}$$

Next we express the indicator function in (4.5) using Green functions. Recall (3.28), a key observation is that the size of $[E^-, E_U]$ is of order $N^{-2/3}$ due to (3.4). As we now use (4.2) and (4.6) in the bulks, the size here is of order 1. So we cannot use the delta approximation function to estimate $\mathcal{X}(E)$. Instead, we use Helffer–Sjöstrand functional calculus. This has been used many times when the window size η takes the form of (4.2), for example, in the proofs of rigidity of eigenvalues in [16], [18], and [33].

For any $0 < E_1,\ E_2 \le \tau^{-1}$, let $f(\lambda) \equiv f_{E_1,E_2,\eta_d}(\lambda)$ be the characteristic function of $[E_1, E_2]$ smoothed on the scale

$$\eta_d := N^{-1-d\varepsilon_0}, \qquad d > 2,$$

where $f = 1$ when $\lambda \in [E_1, E_2]$ and $f = 0$ when $\lambda \in \mathbb{R} \setminus [E_1 - \eta_d, E_2 + \eta_d]$, and

$$|f'| \le C\eta_d^{-1}, \qquad |f''| \le C\eta_d^{-2}, \tag{4.7}$$

for some constant $C > 0$. By Equation (B.12) of [19], with $f_E \equiv f_{E^-,E_U,\eta_d}$, we have

$$f_E(\lambda) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \frac{\mathrm{i}\sigma f_E''(e)\chi(\sigma) + \mathrm{i} f_E(e)\chi'(\sigma) - \sigma f_E'(e)\chi'(\sigma)}{\lambda - e - \mathrm{i}\sigma}\,\mathrm{d}e\,\mathrm{d}\sigma, \tag{4.8}$$

where $\chi(y)$ is a smooth cutoff function with support $[-1, 1]$ and $\chi(y) = 1$ for $|y| \le \frac{1}{2}$ with bounded derivatives. Using a similar argument to that used for Lemma 3.3, we have the following result, whose proof is given in the supplementary material [14].

Lemma 4.4. *Recall the smooth cutoff function q defined in (3.27). Under the assumptions of Lemma 4.3, there exists a $0<\delta<1$ such that*

$$\lim_{N\to\infty} \max_{\delta N_k \le l \le (1-\delta)N_k} \max_{\mu,\nu} \Big| \mathbb{E}^V \theta\Big(\frac{N}{\pi}\int_I \tilde{G}_{\mu\nu}(z)\mathcal{X}(E)\Big)\, \mathrm{d}E - \mathbb{E}^V \theta\Big(\frac{N}{\pi}\int_I \tilde{G}_{\mu\nu}(z) q(\operatorname{Tr} f_E(Q_1))\Big)\, \mathrm{d}E \Big| = 0. \tag{4.9}$$

Finally, we apply the Green function comparison argument, where we will follow the basic approach of Section 3.2 and [24, Section 5]. The key difference is that we will use (4.2) and (4.3).

Lemma 4.5. *Under the assumptions of Lemma 4.4, there exists a $0<\delta<1$ such that*

$$\lim_{N\to\infty} \max_{\delta N_k \le l \le (1-\delta)N_k} \max_{\mu,\nu} [\mathbb{E}^V - \mathbb{E}^G]\theta\Big[\frac{N}{\pi}\int_I \tilde{G}_{\mu\nu}(E+\mathrm{i}\eta) q(\operatorname{Tr} f_E(Q_1))\, \mathrm{d}E\Big] = 0. \tag{4.10}$$

Proof. Recall (4.8). By (2.5), we have

$$\operatorname{Tr} f_E(Q_1) = \frac{N}{2\pi}\int_{\mathbb{R}^2} \big(\mathrm{i}\sigma f_E''(e)\chi(\sigma) + \mathrm{i} f_E(e)\chi'(\sigma) - \sigma f_E'(e)\chi'(\sigma)\big) m_2(e+\mathrm{i}\sigma)\, \mathrm{d}e\, \mathrm{d}\sigma. \tag{4.11}$$

Define $\tilde{\eta}_d := N^{-1-(d+1)\varepsilon_0}$. We can decompose the right-hand side of (4.11) as

$$\begin{aligned} \operatorname{Tr} f_E(Q_1) = {} & \frac{N}{2\pi}\int\int_{\mathbb{R}^2} (\mathrm{i} f_E(e)\chi'(\sigma) - \sigma f_E'(e)\chi'(\sigma)) m_2(e+\mathrm{i}\sigma)\, \mathrm{d}e\, \mathrm{d}\sigma \\ & + \frac{\mathrm{i}N}{2\pi}\int_{|\sigma|>\tilde{\eta}_d} \sigma\chi(\sigma)\int f_E''(e) m_2(e+\mathrm{i}\sigma)\, \mathrm{d}\sigma\, \mathrm{d}e \\ & + \frac{\mathrm{i}N}{2\pi}\int_{-\tilde{\eta}_d}^{\tilde{\eta}_d} \sigma\chi(\sigma)\int f_E''(e) m_2(e+\mathrm{i}\sigma)\, \mathrm{d}\sigma\, \mathrm{d}e. \end{aligned}$$

By (4.3) and (4.7), for some constant $C>0$, with probability $1-N^{-D_1}$, we have

$$\Big|\frac{\mathrm{i}N}{2\pi}\int_{-\tilde{\eta}_d}^{\tilde{\eta}_d} \sigma\chi(\sigma)\int f_E''(e) m_2(e+\mathrm{i}\sigma)\, \mathrm{d}\sigma\, \mathrm{d}e\Big| \le N^{-C\varepsilon_0}. \tag{4.12}$$

Recall (3.35) and (3.38). Similarly to Lemma 3.6, we first drop the diagonal terms. By (4.1), with probability $1-N^{-D_1}$, we have (recall (3.41))

$$\int_I \Big|\frac{N}{\pi}\tilde{G}_{\mu\nu}(E+\mathrm{i}\eta) - x(E)\Big|\, \mathrm{d}E \le N^{-1+C\varepsilon_0}$$

for some constant $C>0$. Hence, by the mean value theorem, we need only prove that

$$\lim_{N\to\infty} \max_{\delta N_k \le l \le (1-\delta)N_k} \max_{\mu,\nu} [\mathbb{E}^V - \mathbb{E}^G]\theta\Big(\int_I x(E) q(\operatorname{Tr} f_E(Q_1))\, \mathrm{d}E\Big) = o(1).$$

Furthermore, by Taylor's expansion, (4.12), and the definition of χ, it suffices to prove that

$$\lim_{N\to\infty} \max_{\delta N_k \le l \le (1-\delta)N_k} \max_{\mu,\nu} [\mathbb{E}^V - \mathbb{E}^G]\theta\Big(\int_I x(E) q(y(E)+\tilde{y}(E))\, \mathrm{d}E\Big) = o(1), \tag{4.13}$$

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

where

$$y(E) := \frac{N}{2\pi} \int_{\mathbb{R}^2} \mathrm{i}\sigma f_E''(e)\chi(\sigma) m_2(e+\mathrm{i}\sigma) \mathbf{1}(|\sigma| \geq \tilde{\eta}_d)\, \mathrm{d}e\, \mathrm{d}\sigma, \tag{4.14}$$

$$\tilde{y}(E) := \frac{N}{2\pi} \int_{\mathbb{R}^2} (\mathrm{i} f_E(e)\chi'(\sigma) - \sigma f_E'(e)\chi'(\sigma)) m_2(e+\mathrm{i}\sigma)\, \mathrm{d}e\, \mathrm{d}\sigma. \tag{4.15}$$

Next we will use the Green function comparison argument to prove (4.13). In the proof of Lemma 3.5, we used the resolvent expansion until an order of four. However, due to the larger bounds in (4.3), we will use the expansion

$$S = R - RVR + (RV)^2 R - (RV)^3 R + (RV)^4 R - (RV)^5 S. \tag{4.16}$$

Recall (3.47) and (3.48). We have

$$\begin{aligned}
&[\mathbb{E}^V - \mathbb{E}^G]\theta\Big(\int_I x(E) q(y(E) + \tilde{y}(E))\, \mathrm{d}E \Big) \\
&= \sum_{\gamma=1}^{\gamma_{\max}} \mathbb{E}\left(\theta\Big(\Big(\int_I x^S q(y^S + \tilde{y}^S) \Big)\Big) - \theta\Big(\Big(\int_I x^T q(y^T + \tilde{y}^T) \Big)\Big) \right).
\end{aligned} \tag{4.17}$$

We still use the same notation $\Delta x(E) := x^S(E) - x^R(E)$. We basically follow the approach of Section 3.2, where the control (3.36) is replaced by (4.3). We first deal with $x(E)$. Let $\Delta x^{(k)}(E)$ denote the summations of the terms in $\Delta x(E)$ containing k numbers of $X^G_{i\mu_1}$. Similarly to the discussion of Lemma 3.7, recalling (3.52), by (1.2) and (4.3), with probability $1 - N^{-D_1}$, we have

$$|\Delta x^{(5)}(E)| \leq N^{-3/2 + C\varepsilon_0}, \qquad M+1 \leq k \neq \mu, \nu \leq M+N.$$

This yields

$$\Delta x(E) = \sum_{p=1}^{4} \Delta x^{(p)}(E) + O(N^{-3/2 + C\varepsilon_0}). \tag{4.18}$$

Let

$$\Delta \tilde{y}(E) = \tilde{y}^S(E) - \tilde{y}^R(E), \qquad \Delta m_2 := m_2^S - m_2^R = \frac{1}{N} \sum_{\mu=M+1}^{M+N} (S_{\mu\mu} - R_{\mu\mu}).$$

We first deal with (4.15). By the definition of χ, we need to restrict $\frac{1}{2} \leq |\sigma| \leq 1$; hence, by (2.17), with probability $1 - N^{-D_1}$, we have

$$\max_{\mu} |G_{\mu\mu}| \leq N^{\varepsilon_1}, \qquad \max_{\mu \neq \nu} |G_{\mu\nu}| \leq N^{-1/2 + \varepsilon_1}. \tag{4.19}$$

By (3.50), (3.51), (4.16), and (4.19), with probability $1 - N^{-D_1}$, we have $|\Delta m_2^{(5)}| \leq N^{-7/2 + 9\varepsilon_1}$. This yields the decomposition

$$\Delta \tilde{y}(E) = \sum_{p=1}^{4} \Delta \tilde{y}^{(p)}(E) + O(N^{-5/2 + C\varepsilon_0}). \tag{4.20}$$

https://doi.org/10.1017/apr.2019.10 Published online by Cambridge University Press

Next we will control (4.14). Define $\Delta y(E) := y^S(E) - y^R(E)$. By (3.50), (3.51) and (4.1), using a similar discussion to that used for Equation (5.22) of [24], with probability $1 - N^{-D_1}$, for $\sigma \geq \tilde{\eta}_d$, we have

$$|\Delta m_2^{(5)}| \leq N^{-5/2+C\varepsilon_0}(N^{-1} + \Lambda_\sigma^2), \tag{4.21}$$

where $\Lambda_\sigma := \sup_{|e| \leq \tau^{-1}} \max_{\mu \neq \nu} |G_{\mu\nu}(e + \mathrm{i}\sigma)|$, recalling that $\mu, \nu \in \mathcal{I}_2$. In order to estimate $\Delta y(E)$, we integrate (4.14) by parts, first in e then in σ. By Equation (5.24) of [24], with probability $1 - N^{-D_1}$, we have

$$\begin{aligned} \Big| \frac{N}{2\pi} \int_{\mathbb{R}^2} &\mathrm{i}\sigma f_E''(e) \chi(\sigma) \Delta^{(5)} m_2(e + \mathrm{i}\sigma) \mathbf{1}(|\sigma| \geq \tilde{\eta}_d) \, \mathrm{d}e \, \mathrm{d}\sigma \Big| \\ &\leq CN \Big| \int f_E'(e) \tilde{\eta}_d \Delta m_2^{(5)}(e + \mathrm{i}\tilde{\eta}_d) \, \mathrm{d}e \Big| + CN \Big| \int f_E'(e) \, \mathrm{d}e \int_{\tilde{\eta}_d}^{\infty} \chi'(\sigma) \sigma \, \Delta m_2^{(5)(e+\mathrm{i}\sigma)} \, \mathrm{d}\sigma \Big| \\ &+ CN \Big| \int f_E'(e) \, \mathrm{d}e \int_{\tilde{\eta}_d}^{\infty} \chi(\sigma) \Delta m_2^{(5)}(e + \mathrm{i}\sigma) \, \mathrm{d}\sigma \Big|. \end{aligned} \tag{4.22}$$

By (4.21), with probability $1 - N^{-D_1}$, the first two items of (4.22) can be easily bounded by $N^{-5/2+C\varepsilon_0}$. For the last item, by (4.21), (4.1), and a similar discussion to the equation below [24, Equation (5.24)], it can be bounded by

$$CN \int_{\tilde{\eta}_d}^{1} \Big(\frac{1}{N\sigma} + \frac{1}{(N\sigma)^2} + \frac{1}{N} \Big) N^{-5/2+C\varepsilon_0} \leq N^{-5/2+C\varepsilon_0}.$$

Hence, with probability $1 - N^{-D_1}$, we have the decomposition

$$\Delta y(E) = \sum_{p=1}^{4} \Delta y^{(p)}(E) + O(N^{-5/2+C\varepsilon_0}). \tag{4.23}$$

Similarly to the discussion of (4.18), (4.20), and (4.23), it is easy to check that, with probability $1 - N^{-D_1}$, we have

$$\begin{aligned} \int_I |\Delta x^{(p)}(E)| \, \mathrm{d}E \leq N^{-p/2+C\varepsilon_0}, \qquad |\Delta \tilde{y}^{(p)}(E)| \leq N^{-p/2+C\varepsilon_0}, \\ |\Delta y^{(p)}(E)| \leq N^{-p/2+C\varepsilon_0}, \end{aligned} \tag{4.24}$$

where $p = 1, 2, 3, 4$ and $C > 0$ is some constant. Furthermore, by (4.1), with probability $1 - N^{D_1}$, we have

$$\int_I |x(E)| \, \mathrm{d}E \leq N^{C\varepsilon_0}. \tag{4.25}$$

Due to the similarity of (4.20) and (4.23), letting $\bar{y} = y + \tilde{y}$, we have

$$\Delta \bar{y} = \sum_{p=1}^{4} \Delta \bar{y}^{(p)}(E) + O(N^{-5/2+C\varepsilon_0}). \tag{4.26}$$

By (4.24), (4.26), and Taylor's expansion, we have

$$q(\bar{y}^S) = q(\bar{y}^R) + q'(\bar{y}^R)\Big(\sum_{p=1}^{4} \Delta\bar{y}^{(p)}(E)\Big) + \frac{1}{2}q''(\bar{y}^R)\Big(\sum_{p=1}^{3} \Delta\bar{y}^{(p)}(E)\Big)^2 + \frac{1}{6}q^{(3)}(\bar{y}^R)\Big(\sum_{p=1}^{2} \Delta\bar{y}^{(p)}(E)\Big)^3 + \frac{1}{24}q^{(4)}(\bar{y}^R)\Big(\Delta\bar{y}^{(1)}(E)\Big)^4 + o(N^{-2}). \tag{4.27}$$

By (4.4), we have

$$\theta\Big(\int_I x^S q(\bar{y}^S)\,\mathrm{d}E\Big) - \theta\Big(\int_I x^R q(\bar{y}^R)\,\mathrm{d}E\Big) = \sum_{s=1}^{4} \frac{1}{s!}\theta^{(s)}\Big(\int_I x^R q(\bar{y}^R)\,\mathrm{d}E\Big)\Big[\int_I x^S q(\bar{y}^S)\,\mathrm{d}E - \int_I x^R q(\bar{y}^R) dE\Big]^s + o(N^{-2}). \tag{4.28}$$

Inserting $x^S = x^R + \sum_{p=1}^4 \Delta x^{(p)}$ and (4.27) into (4.28), using the partial expectation argument as in Section 3.2, by (4.4), (4.24), and (4.25), we find that there exists a random variable B that depends on the randomness only through O and the first four moments of $X^G_{i\mu_1}$ such that

$$\mathbb{E}\theta\Big(\int_I x^S q(y+\tilde{y})^S\,\mathrm{d}E\Big) - \mathbb{E}\theta\Big(\int_I x^R q(y+\tilde{y})^R\,\mathrm{d}E\Big) = B + o(N^{-2}).$$

Hence, together with (4.17), this proves (4.13), which implies (4.10). This completes our proof. □

Acknowledgements

I am very grateful to Jeremy Quastel and Bálint Virág for many valuable insights and helpful suggestions, which have significantly improved the paper. I would like to thank my friend Fan Yang for many useful discussions and pointing out some references, especially [39]. I also want to thank two anonymous referees, the Associate Editor, and the Editor for their many helpful comments.

References

[1] Abbe, E., Fan, J., Wang, K. and Zhong. Y. (2017). Entrywise eigenvector analysis of random matrices with low expected rank. Preprint. Available at https://arxiv.org/abs/1709.09565.

[2] Bai, Z. D. and Silverstein, J. W. (2010). Spectral Analysis of Large Dimensional Random Matrices, 2nd edn. Springer, New York.

[3] Bai, Z. D., Miao, B. Q. and Pan, G. M. (2007). On asymptotics of eigenvectors of large sample covariance matrix. *Ann. Prob.* **35**, 1532–1572.

[4] Baik, J., Ben Arous, G. and Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Prob.* **33**, 1643–1697.

[5] Bao, Z., Pan, G. and Zhou, W. (2015). Universality for the largest eigenvalue of sample covariance matrices with general population. *Ann. Statist.* **43**, 382–421.

[6] Benaych-Georges, F. and Nadakuditi, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227**, 494–521.

[7] Benaych-Georges, F., Guionnet, A. and Maida, M. (2011). Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electron. J. Prob.* **16**, 1621–1662.

[8] BLOEMENDAL, A., KNOWLES, A., YAU, H.-T. AND YIN, J. (2016). On the principal components of sample covariance matrices. *Prob. Theory Relat. Fields* **164**, 459–552.
[9] BLOEMENDAL, A. *et al.* (2014). Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Prob.* **19**, pp. 53.
[10] BOURGADE, P. AND YAU, H.-T. (2017). The eigenvector moment flow and local quantum unique ergodicity. *Commun. Math. Phys.* **350**, 231–278.
[11] CAI, T. T., REN, Z. AND ZHOU, H. H. (2016). Estimating structured high-dimensional covariance and precision matrices: optimal rates and adaptive estimation. *Electron. J. Statist.* **10**, 1–59.
[12] CAPITAINE, M., DONATI-MARTIN, C. AND FÉRAL, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations. *Ann. Prob.* **37**, 1–47.
[13] DING, X. (2017). High dimensional deformed rectangular matrices with applications in matrix denoising. Preprint. Available at https://arxiv.org/abs/1702.06975.
[14] DING, X. (2019). Singular vector distribution of sample covariance matrices. Supplementary material. Available at http://doi.org/10.1017/apr.2019.10.
[15] DING, X. AND SUN, Q. (2018). Modified Multidimensional Scaling and High Dimensional Clustering. Available at https://arxiv.org/abs/1810.10172.
[16] DING, X. AND YANG, F. (2018). A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices. *Ann. Appl. Prob.* **28**, 1679–1738.
[17] EL KAROUI, N. (2007). Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Prob.* **35**, 663–714.
[18] ERDŐS, L., YAU, H.-T. AND YIN, J. (2012). Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.* **229**, 1435–1515.
[19] ERDŐS, L., RAMIREZ, J. A., SCHLEIN, B. AND YAU, H.-T. (2010). Universality of sine-kernel for Wigner matrices with a small Gaussian perbubation. *Electron. J. Prob.* **15**, 526–603.
[20] FAN, J. AND ZHONG, Y. (2018). Optimal subspace estimation using overidentifying vectors via generalized method of moments Available at https://arxiv.org/abs/1805.02826.
[21] FAN, J., WANG, W. AND ZHONG, Y. (2017). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *J. Machine Learning Res.* **18**, pp. 42.
[22] GOLUB, G. H. AND VAN LOAN, C. F. (2013). *Matrix Computations*, 4th edn. John Hopkins University Press, Baltimore, MD.
[23] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29**, 295–327.
[24] KNOWLES, A. AND YIN, J. (2013). Eigenvector distribution of Wigner matrices. *Prob. Theory Relat. Fields* **155**, 543–582.
[25] KNOWLES, A. AND YIN, J. (2014). The outliers of a deformed Wigner matrix. *Ann. Prob.* **42**, 1980–2031.
[26] KNOWLES, A. AND YIN, J. (2017). Anisotropic local laws for random matrices. *Prob. Theory Relat. Fields* **169**, 257–352.
[27] LEDOIT, O. AND PÉCHÉ, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Prob. Theory Relat. Fields* **151**, 233–264.
[28] LEE, J. O. AND SCHNELLI, K. (2016). Tracy-Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *Ann. Appl. Prob.* **26**, 3786–3839.
[29] LI, G., TANG, M., CHARON, N. AND PRIEBE, C. E. (2018). A central limit theorem for classical multidimensional scaling. Available at https://arxiv.org/abs/1804.00631.
[30] MARČENKO, V. A. AND PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sbornik* **1**, 457.
[31] O'ROURKE, S., VU, V. AND WANG, K. (2016). Eigenvectors of random matrices: a survey. *J. Combinatorial Theory Ser A* **144**, 361–442.
[32] PILLAI, N. S. AND YIN, J. (2012). Edge universality of correlation matrices. *Ann. Statist.* **40**, 1737–1763.
[33] PILLAI, N. S. AND YIN, J. (2014). Universality of covariance matrices. *Ann. Appl. Prob.* **24**, 935–1001.
[34] SILVERSTEIN, J. W. (1984). Some limit theorems on the eigenvectors of large-dimensional sample covariance matrices. *J. Multivariate Anal.* **15**, 295–324.
[35] SILVERSTEIN, J. W. (2009). The Stieltjes transform and its role in eigenvalue behavior of large dimensional random matrices. in *Random Matrix Theory and its Applications*, World Scientific, Hackensack,
[36] SILVERSTEIN, J. W. AND CHOI, S.-I. (1995). Analysis of the limiting spectral distribution of large dimensional random matrices. *J. Multivariate Anal.* **54**, 295–309.
[37] TAO, T. AND VU, V. (2012). Random matrices: universal properties of eigenvectors. *Random Matrices Theory Appl.* **1**, pp. 27.

[38] TRACY, C. A. AND WIDOM, H. (1996). On orthogonal and symplectic matrix ensembles. *Commun. Math. Phys.* **177**, 727–754.
[39] XI, H., YANG, F. AND YIN, J. (2017). Local circular law for the product of a deterministic matrix with a random matrix. *Electron. J. Prob.* **22**, pp. 77.
[40] YAO, J., ZHENG, S. AND BAI, Z. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. Cambridge University Press.
[41] ZHONG, Y. AND BOUMAL, N. (2018). Near-optimal Bounds For Phase Synchronization. *SIAM J. Optimization* **28**, 989–1016.